Journal of
Experimental Psychology

VOL. 23, NO. 6 DECEMBER, 1938

ANISEIKONIA AS RELATED TO DISABILITY IN READING[1]

BY WALTER F. DEARBORN AND IRVING H. ANDERSON

Harvard University

It is well known that some intelligent children (not to speak of unintelligent children) are handicapped by distinct reading disabilities which block progress in their studies. Many of these cases are due to the accumulative effect of such factors as large classes and the resulting lack of individual attention or of neglect on the part of teachers, lack of interest, or of 'readiness' for reading on the part of the pupil, unfortunate conditioning of emotional reactions in the home and in school, and so forth. In a considerable number of these cases and in other cases in which the above factors are not conspicuously present, certain structural conditions appear to be the primary or at least initial causative factors. Ocular defects and faulty hearing are among the more important of these conditions. This, of course, does not mean that every child with defective vision or hearing will necessarily be retarded in reading. Errors of refraction and muscular imbalance, especially if the fusional reserves are high, are relatively insignificant diagnostic signs when studied independently of such other factors as are mentioned above.

[1] This investigation has been made possible by a subvention from the Carnegie Corporation, granted on the recommendation of the Carnegie Foundation for the Advancement of Teaching. Acknowledgment is herewith made also to Mr. Henry A. Imus of the Research Division of the Dartmouth Eye Institute, Dartmouth Medical School, who has served as consultant on many of the technical aspects of this paper.

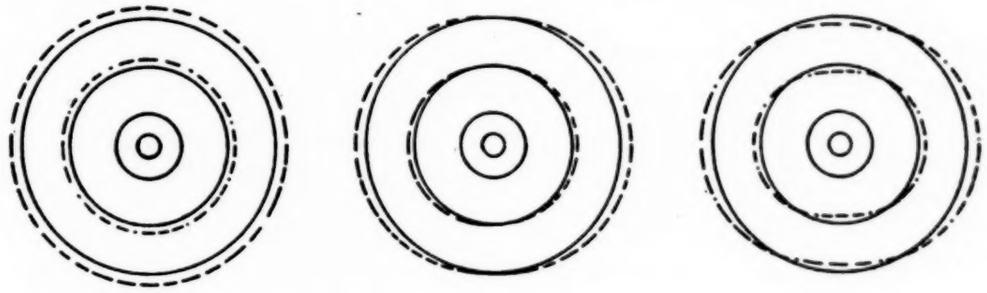
Similarly, impairment of hearing, unless unusually grave or present in a situation where phonetic methods are chiefly employed in the teaching of reading, may be but of slight importance. However, if the defects of vision or hearing are rather pronounced and the child's ability to compensate for them is low, and especially if the teacher does nothing to modify her method of instruction so as to minimize the effects of these handicaps, they may result in failure in the early attempts at learning to read. The most likely hypothesis, then, of the way in which structural defects may operate to produce retardation in reading is that they are in the nature of initial handicaps, more or less readily surmounted by the great majority of children, but which trip the faltering steps of a few individuals either because the amount of impairment is in excess of that which they are prepared to overcome or because the structural conditions occur in combination with each other or with other barriers to learning, such as those above indicated.

The present article offers evidence for the addition of another structural condition to the list of possible causative factors of disability in reading. Aniseikonia [2] (*ana*—negative; *iso*—equal; *eikon*—image) is a condition of the eyes in which the ocular images [3] are unequal either in size or shape, so that during binocular vision conflicting excitations arrive at the visual centers in the occipital cortex and present difficulties in the ability of the individual to fuse the incongruent images aroused. Of the various types of aniseikonia, the most common are: (1) overall differences, in which the image formed

[2] Ames, A., Jr., *Aniseikonia. The Eye and Its Diseases* (Edited by C. Bevens), Phila., W. B. Saunders Co., 1936.

[3] The term ocular image is used to describe the impression which reaches consciousness through the vision of one eye. It is determined not only by the properties of the dioptric image that is formed on the retina of each eye, but also by the modifications imposed upon that image by the anatomical properties and physiological processes by which this optical image is transmitted to the higher visual centers. Thus in aniseikonia there is a difference as between the two eyes in the perception of size and shape. The subject is ordinarily not aware of these differences either because they are too small or because he has learned to neglect them or to adjust himself to them. The effect of aniseikonia can be demonstrated by placing before the eyes aniseikonic lenses. The subject in such demonstrations gets the effect or what may have been the effect before he had learned to adjust to these differences in size and shape.

in one eye is relatively larger or smaller in both the vertical and horizontal meridians than that formed in the other eye;[4] (2) simple meridional differences, in which the image formed in one eye is larger or smaller than that of the other in only one of these meridians; and (3) compound meridional differences, in which the image formed in one eye is larger in the one meridian but smaller in the opposite meridian as compared with the image of the other eye in the corresponding meridians. These types of aniseikonia are illustrated on Fig. 1. The



1 2 3

FIG. 1. Schematic representation of three common types of aniseikonia: 1, 2, and 3 represent overall, simple meridional, and compound meridional differences in the size and shape of ocular images, respectively.

dotted lines represent the image formed in one eye, the solid lines that formed in the other. There may, of course, be combinations and variations of the three types portrayed, as, for example, in the case in which an overall difference is combined with a meridional one.

Differences in the size and shape of ocular images may affect binocular vision in two main ways. First, the disparity of the images relative to corresponding points[5] and areas on the two retinas is atypical. As a result, the binocular stereoscopic interpretation of the positioning of planes and surfaces is disturbed.[6] Second, the inequality of the images presents difficulties in fusion, which may be manifested by

[4] Although size differences may also exist in the oblique axes, it has not been practicable clinically to measure aniseikonia in these directions.

[5] Ames, A., Jr., Ogle, K. N., and Gliddon, G. H., Corresponding retinal points, the horopter and size and shape of ocular images, *J. Optical Soc. Amer.*, 1932, **22**, 538–574, 575–632.

[6] Ames, A., Jr., Aniseikonia, a factor in the functioning of vision, *Amer. J. Ophthal.*, 1935, **18**, 1014–1019.

symptoms of visual discomfort, photophobia, headache, and other signs of a neurological nature,[7] or may even prevent fusion entirely if the difference in relative size exceeds five per cent.[8]

The way in which these disturbances may in turn specifically affect reading is as follows: First, because of the interference with the ability to fuse,[9] the aniseikonia may produce either ocular or general fatigue, or both, and thereby raise the threshold of effective motivation and decrease the length of the periods of concentration. This inability to persevere, especially while reading for the purposes of study, is a common complaint of aniseikonic patients. After relatively brief periods of reading, many of these cases find themselves perusing pages mechanically without any recollection of the material covered. Second, aniseikonia may interfere directly with the visual processes during reading. Its effect in this connection, however, may begin only when the reader tries to increase his perceptual span, that is, when he attempts to read in larger units than single words or syllables. Reading, like efficient driving of an automobile in traffic, demands a very skillful use of peripheral vision. It is with this peripheral view of the line of print, which is so requisite for the proper spacing of the fixation pauses, that aniseikonia may interfere. This follows because the incongruent images falling on the periphery of each eye would present different space cues for the movements of the eyes as they are shifted from one point within the line to the next. Furthermore, since the paper or book may appear tilted because of the aniseikonia, slight readjustments of the accommodation and convergence may be necessary. The need of having to make such readjustments might, in turn, result in an excessive number of fixations and regressions of long and variable duration. If so, it would be difficult for the person to read as well as he otherwise might.

[7] Bielschowsky, A., Functional disturbances of the eyes, *Arch. Ophthal.*, 1936, **15**, 589–603.

[8] Bielschowsky, A., Congenital and acquired deficiencies of fusion, *Amer. J. Ophthal.*, 1935, **18**, 925–937.

[9] Bielschowsky, A., Lectures on motor anomalies of the eyes, *Arch. Ophthal.*, 1934, **12**, 805–818.

Experimental Procedure

Apparatus.—Aniseikonia is not discovered spontaneously by the subject, nor can it be determined by the procedures ordinarily employed in examining the eyes for errors of refraction and motor anomalies. An accurate instrument, known as the Ophthalmo-Eikonometer, has been designed and built to measure inequalities in the size and shape of ocular images, and a careful, scientific technique has been developed to obtain reliable data. The instrument and technique originated in the Dartmouth Eye Institute of the Dartmouth Medical School.[10]

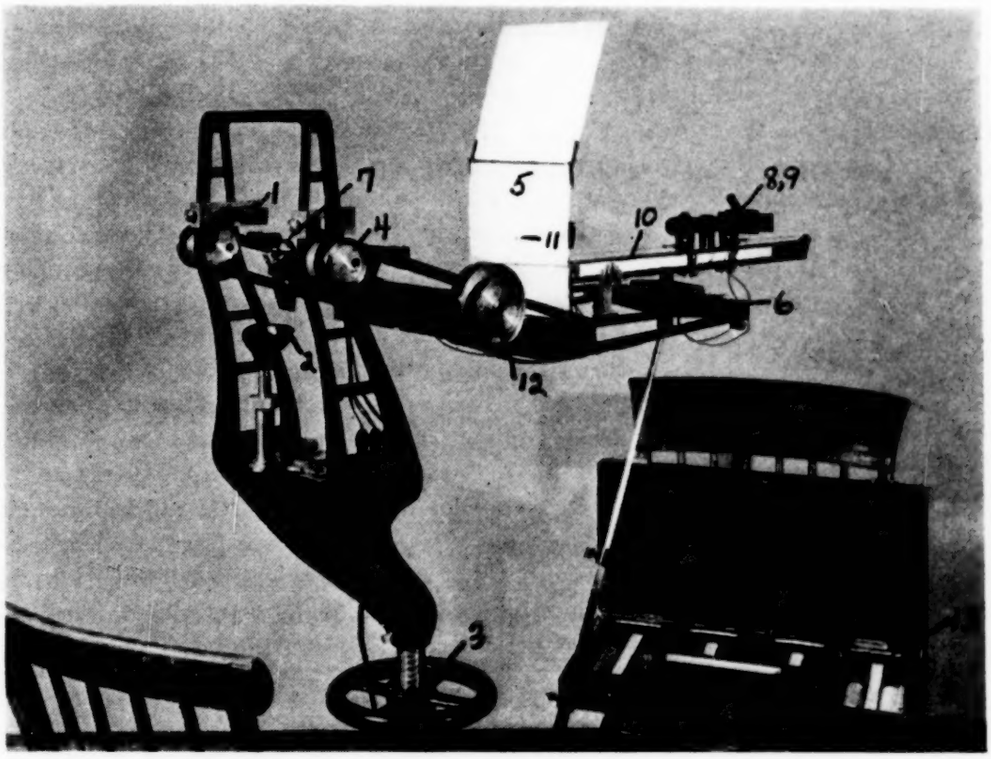


Fig. 2. Instrument for measuring a difference in the size and shape of ocular images (the Ophthalmo-Eikonometer). (1) Adjustable head rest, (2) chin rest, (3) adjustable base, (4) corneal aligning devices, (5) target, (6) double arms, (7) spherical power distortionless prisms, (8, 9) illuminated letter chart within an optical system, (10) track, (11) aperture, (12) focussing wheel, (13) test set of aniseikonic lenses.

The model used in this study (Fig. 2) was designed for the Psycho-Educational Clinic of Harvard University by the above Department in 1931. Although other models of the instrument have been described elsewhere,[11] a brief description of the essential features of this model is necessary.

The instrument consists of the following: an adjustable head rest (1), chin rest (2), supported by an adjustable base (3), and corneal aligning devices (4), by means of which the eyes are placed in the instrument so that they are equidistant from the plane

[10] Ames, A., Jr., Gliddon, G. H., and Ogle, K. N., Size and shape of ocular images, *Arch. Ophthal.*, 1932, **7**, 576–597.

[11] Hughes, W. L., Some clinical observations on aniseikonia, *Amer. J. Ophthal.*, 1935, **18**, 716.

of the target (5), and become an integral part of the optical system. The double arms (6), extending to the right on the figure, are pivoted about the centers of rotation of the eyes, so that the target may be placed in position for distant vision (as shown), or for near vision (tipped down through an arc of 20°). Especially designed spherical power distortionless prisms (7) may be swung into position before the eyes so that the target is projected optically to a distance of 20 feet, which is the usual distance for testing the eyes. These prisms are not used, of course, when the eyes are being tested at the reading distance. An illuminated letter chart (8), within an optical system (9), is free to travel upon the long track (10), which extends beyond the double arms. The chart can be observed and focussed by the subject monocularly through the aperture (11), in the center of the target. This adjustment is made while the eyes are fixating a fine cross-hair suspended in the center of the target opening. The focussing is done by turning the large wheel (12), on the right of the double arm, which is connected by cord and pulleys to the traveling system. In this manner the point in space to which the retina of each eye is conjugate may be determined while the normal convergence and accommodation of both eyes are controlled. The position of these points relative to the position of the target is a measure of the refractive errors of the two eyes for the particular test distance. The refractive corrections are varied until the refractive conditions of both eyes are normal and equal, that is, the points in space conjugate to the retinas of both eyes are normal to the target and are equidistant.[12] It is essential to comply with these conditions before beginning the test for aniseikonia.

In order to test for aniseikonia, it is necessary to present targets to the two eyes which have identical designs to demand fusion, and dissimilar designs to permit differentiation of the two ocular images. This is accomplished by the use of a target (5), which has a black center seen by both eyes, and four tiny apertures, 0.18 mm in diameter, placed 4° equidistant from the center at the middle of each peripheral edge and illuminated from behind in such a way that they appear as bright points of light to the right eye and as black dots on a white background to the left eye. Thus, when the subject fixates the identical designs at the center and fuses them, he sees both a light

[12] Refractive errors are corrected by trial lenses which are placed before the eyes in special holders attached to the instrument. The final correction for the particular test distance is permitted to remain in place during the course of the examination for aniseikonia. If the refractive errors of the two eyes differ markedly and are fully corrected, there will result an induced difference in the size of the two retinal images which may be measured and corrected in the same way that aniseikonia, which does not have this optical basis, is measured and corrected. Because of the difference in image size which an anisometropia of high degree may induce (and thus produce various symptoms of ocular distress) it has not been deemed advisable heretofore to give such cases the full correction. Instead, the practice has been to give each eye a partial correction in order to reduce the difference in power as between the lenses prescribed. Sometimes the eye having the poorer vision is not corrected at all. In the light of the recent work on aniseikonia, however, it is possible now to give the full refractive correction because whatever difference in image size is induced may be corrected by means of aniseikonic lenses. It may well be that in some cases of poor reading an aniseikonic correction is effective mainly because it permits a full correction of refractive errors. This is not to be construed to mean, however, that all poor readers who profit by a correction for aniseikonia are helped on this basis alone. Cases of anisometropia sufficient to cause discomfort are comparatively rare, even among poor readers.

and a dot at each of four cardinal points, namely, at the peripheries at the top, bottom, left, and right. If the subject has no significant amount of aniseikonia, he sees the lights and dots superimposed, or symmetrically situated relative to each other; that is, the lights may be seen displaced in the same direction and by the same amount due to fixation disparity. If the ocular images are unequal in size and shape, he reports the lights and dots situated asymmetrically relative to each other. For example, if the vertical meridian is the one being measured, he may report the light at the top as being higher than the dot, and the light at the bottom as being lower, in which case the image of the eye seeing the lights would be larger in that meridian. If in this particular case the aniseikonia was of the simple meridional type, the light and dot at each of the poles of the horizontal meridian would be reported as being superimposed or situated symmetrically with respect to each other. Otherwise, the lights would be outside of the dots or toward the center from them, depending upon whether the aniseikonia was of the overall or compound meridional type, respectively.)

By means of a set[13] of overall and meridional zero verging power lenses,[14] which have the effect of magnifying one ocular image relative to the other in all meridians or in one meridian alone, respectively, without changing the refractive condition of the eyes, it is possible to determine the amount of aniseikonia within an accuracy of from 0.25 to 0.50 percent. All that need be done is to determine the lens combination which will equalize the two ocular images in all meridians. The clinical procedure is to increase the size of the smaller image in the meridian being measured by 0.25 percent steps until the point is reached where the subject first reports that the lights and dots are superimposed or situated symmetrically relative to each other. Note is made of the strength of the lens immediately below the one which produces this effect. The clinician then continues to apply additional lens magnification in the same steps until the point is reached where the subject first reports a reversal of the original position of the lights relative to the dots. Note is made of the strength of the lens which produces this effect. The two lenses noted establish the lower and upper limits of the subject's range of amplitude. The true correction is represented by the strength of the lens which lies at the midpoint of this range; that is, the lens which, when increased or decreased by the same amount, just perceptibly reverses the position of the lights relative to the dots is the correct one. The opposite meridian is measured and corrected in the same way, the final correction being a combination of the measurements in the two meridians. The results can be verified by repeated tests at the same sitting or at a later time.

Subjects.—Two groups of subjects were used in this study. The experimental group consisted of the 100 most severe cases of reading disability who had been measured for aniseikonia in the laboratory of the Psycho-Educational Clinic during the period extending from October, 1932, to May, 1938.[15] The amount of retardation in reading was determined by computing the ratio of the difference which existed in the direction of a higher mental than reading age score. Except in the case of two adults,[16] mental

[13] The set of aniseikonic lenses used in this study was loaned to the Psycho-Educational Clinic by the American Optical Company, Southbridge, Mass. (13, Fig. 2).

[14] Ogle, K. N., The correction of aniseikonia with ophthalmic lenses, *J. Optical Soc. Amer.*, 1936, **26**, 323–337.

[15] Grateful acknowledgment is made to Mr. Forrest D. Comfort for his assistance in examining for aniseikonia some of the subjects included in this study.

[16] In the case of these two persons, scores on neither of the above two tests were available. They were included in this study on the basis of a discrepancy between

age was measured either by the Dearborn Group Tests of Mental Ability, Examination C, or by the Stanford Revision of the Binet-Simon Tests, depending upon the one which was administered when the case was examined in the Clinic. When both tests were given, scores on the former were always used because it tends by reason of its partly non-verbal character to be a somewhat more valid measure of the mental ability of reading disability cases than the latter, which is more linguistic in type, especially at the levels at which we needed to use it, namely, from about the age of nine to the adult level. Results on the Dearborn Tests were available on 82 of the 100 subjects. As a group, the reading disability cases tended to be more or less normally distributed as to mental ability, their intelligence quotients ranging from 80 to 170, with an average of 110 and a standard deviation of 9.

Reading ability was measured by a number of tests. Whenever possible, the New Stanford Reading Test was employed because it was the one on which scores were likely to be available on the largest number of cases. Reading age was determined by scores on the paragraph meaning part of this test in the case of 57 of the 100 subjects. When scores were not available on the Stanford Test either because it was not given at the same time that the aniseikonic and mental ability measurements were made or because it was not given at all, scores which were available on the tests next most frequently used in the Clinic were employed to estimate the reading achievement of our subjects. If no reading test scores were reported in the record of a case in the Clinic files, he was, of course, not included in the clinical group of this study, even though he had had the test for aniseikonia. Of the other tests used, the Shank Tests of Reading Comprehension, the Burgess Scale for Measuring Ability in Silent Reading, the Gates Silent Reading Test, and the proper form of the Haggerty Reading Examination were employed in the case of 19, 7, 6, and 5 subjects, respectively. The Chapman-Cook Speed of Reading Test, the Chapman Unspeeded Reading Comprehension Test, and the Nelson-Denny Reading Test were each used in the case of two subjects. It should be repeated that in cases where scores on a more frequently used test were not available the test next most frequently used was always selected to measure reading age. No attempt was made to select tests which indicated the greatest amount of retardation. That the same tests were not available on all cases is an accident of the procedure ordinarily employed in the conduct of the Psycho-Educational Clinic. No attempt is made in the Clinic to give all cases the same battery of tests. Tests are always selected that best meet individual needs. Consequently, the tests chosen vary from case to case, depending upon the age of the child, his abilities, both general and special, the amount of testing time, and a host of other factors.

Excluding the two adult cases above indicated, the amount of difference between mental and reading age of the experimental group varied from eight months in the case of a ten year old boy to slightly more than five years in the case of a sixteen year old boy. The amount of discrepancy between these scores, on the average, was two years and ten months, the standard deviation of the distribution being one year and four months. The specific nature of the reading disability which these cases presented is amply confirmed by an examination of their scholastic histories. Uniformly, they tended to present the typical frustrations which otherwise intelligent children frequently encounter by their failure to read well enough even to cover minimum assign-

---

percentile ranks on the Henmon-Nelson Tests of Mental Ability and the Nelson-Denny Reading Test. In one case there was a difference of 40 percentiles between the two scores, in the other, a difference of over 50. In both cases, of course, the mental ability rating was higher.

ments in a reasonable amount of time. The fact that they were referred to the Clinic, in the first place, for specialized analysis and guidance indicates how critical their condition must have been.

The sources of the control cases used in this study were an unselected group of adults and public school children who had previously been called into the Clinic especially for measurements of aniseikonia, regardless of their ability to read, and a group of clinical cases who had been referred to the Clinic because of scholastic failures and other problems, but who upon examination proved not to have distinct reading disabilities, that is, their reading and mental ages were approximately the same. Of the number of subjects thus available, the 100 who could best be matched for chronological age with the experimental cases were chosen as the control group. The number of subjects finally selected from each of the above two sources was 80 and 20, respectively.

TABLE 1

DISTRIBUTION OF EXPERIMENTAL AND CONTROL CASES ACCORDING TO CHRONOLOGICAL AGE TO NEAREST BIRTHDAY

Chronological Age	Frequency	
	Experimental	Control
9	8	0
10	7	2
11	11	27
12	6	22
13	10	13
14	13	6
15	14	5
16	7	3
17	10	1
18 or over	14	21
Totals	100	100

As is shown by the frequency distributions on Table 1, it was not possible, with the data at hand, to match the groups very closely. The averages, 14.12 and 14.11 years, were practically identical, but the group variances were not correspondingly close, the standard deviations of the two distributions being 3.2 and 3.8 years, respectively, for the experimental and control groups. The fact that the two distributions do not conform more closely than shown is of negligible consequence, however, inasmuch as chronological age and amount of aniseikonia are essentially unrelated, the correlation between the two measures being .13 ± .07.[17] It should be noted that the experimental cases are fairly evenly distributed over the age range used. That there are no cases below the age of nine, in either group, is explained by the fact that it is difficult to obtain reliable measurements of aniseikonia on children as young as that, unless they have mental ages of ten or over. No attempt was made to match the groups on the basis of mental age, for there is no evidence at all to show that mental ability, as such, is related to amount of aniseikonia.

[17] This correlation was computed on 100 unselected cases who ranged in age from nine to twenty-three years and who had aniseikonia in amounts ranging from 0.25 to 7.00 percent.

Results

Incidence of Aniseikonia in Experimental and Control Groups.—The distributions shown on Table 2 present a comparison of the frequency with which aniseikonia of various amounts occurred among the experimental and control cases.

TABLE 2

Distribution of Experimental and Control Cases According to Amount of Aniseikonia in Percent

Aniseikonia in Percent	Frequency			
	Experimental		Control	
0.00	12		17	
0.25	9		27	
0.50	17	49%	21	77%
0.75	11		12	
1.00	9		6	
1.25	5		8	
1.50	8		2	
1.75	5		2	
2.00	6		2	
2.25	4		1	
2.50	3	51%	0	23%
2.75	0		2	
3.00	6		0	
3.25	0		0	
3.50	0		0	
3.75	1		0	
4.00	2		0	
Over 4.00	2		0	
Totals	100		100	

Except in the case of seven subjects in each group, on whom aniseikonic findings were available for distant vision only, the comparison of the two groups is made on the basis of the results for the test at reading distance. The amount of aniseikonia is stated as the difference in the size of the ocular image of one eye relative to the size of that of the other, and is expressed in percent. For example, an aniseikonia of one percent means that the ocular image of one eye is larger (or smaller) than that of the other by one percent, or that a one percent aniseikonic lens placed before the eye having the smaller image will equalize the size of the two ocular images.

Perhaps it will be best to begin our analysis of the results by comparing the number of subjects in each group who had aniseikonia in amounts of 1.00 percent or over, for this is approximately the amount which previous investigators have reported as being necessary to produce visual disturbances.[18, 19, 20, 21, 22] This does not necessarily mean, however, that aniseikonia of less than 1.00 percent is never clinically significant. Since individuals differ widely in their tolerance for this defect, it may well be that in certain cases an aniseikonia of 0.75 percent, or even one of 0.50 percent, may need to be corrected. In general, however, it is customary to ignore amounts less than 1.00 percent. If we proceed on the latter basis, then, and inspect the distributions shown on Table 2, it will be seen that a relatively larger proportion of the experimental than of the control cases had aniseikonia in amounts of 1.00 percent or more, the percentages in the case of the two groups being 51 and 23, respectively. The difference between these percentages is 4.31 times the standard error of the difference. When expressed in terms of probability, this means that the chances are 99.9+in 100 that the difference is a true difference.

Not only was the frequency with which significant amounts of aniseikonia occurred greater in the experimental than in the control group, but of the cases in both groups with aniseikonia of 1.00 percent or more those in the former tended also to have the larger amounts. Consider Fig. 3. It shows graphically the data of Table 2. As may be noted, the amounts of aniseikonia expressed in percent are plotted along the abscissa and the frequency of occurrence of these amounts along the ordinate. The preponderance of cases of reading disability in the really high brackets of aniseikonia is shown strikingly

[18] Ames, A., Jr., Gliddon, G. H., and Ogle, K. N., *op. cit.*

[19] Hughes, W. L., *op. cit.*

[20] Carleton, E. H., and Madigan, L. F., Size and shape of ocular images, *Arch. Ophthalm.*, 1932, 7, 720–738.

[21] Madigan, L. F., and Carleton, E. H., A clinical report of the correction of differences in the size and shape of ocular images, *Ann. Distinguished Service Found. Optometry*, 1936, 1, 83.

[22] Doane, H. C., The clinical significance of differences in the relative size and shape of ocular images, *Trans. Amer. Acad. Optometry*, 1933, 8, 123–124.

on this figure. We need only to compare the range of aniseikonia presented by each of the groups to detect that. Of all subjects with significant amounts of aniseikonia, those in the experimental group had amounts ranging from 1.00 to 7.00 percent, whereas those in the control group were clustered within the range from 1.00 to 2.75 percent. Eleven of the reading disability cases had amounts of aniseikonia in excess of 2.75 percent. Two of these cases had 5.00 and 7.00 percent of aniseikonia. Previous research has shown that amounts

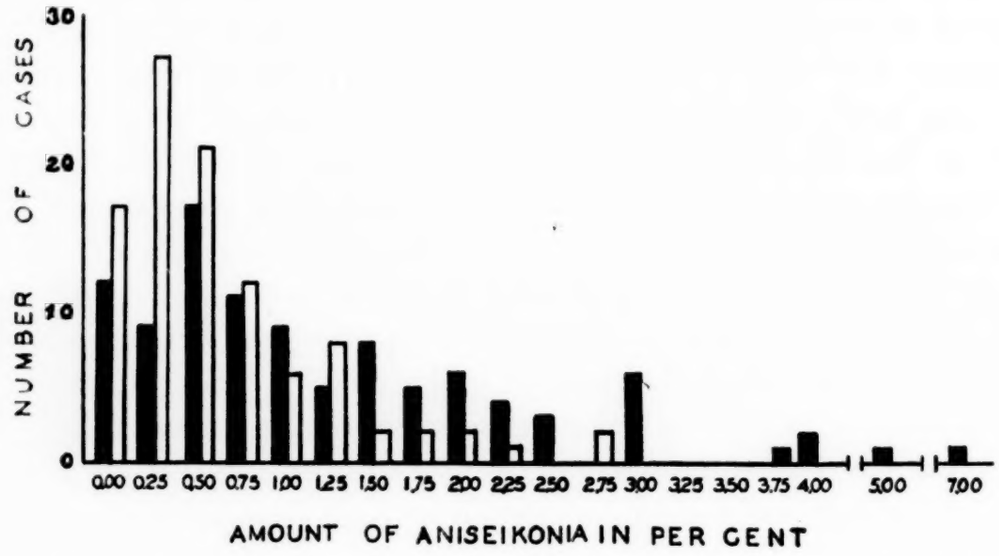


FIG. 3. Comparison of experimental and control groups for incidence of aniseikonia. Solid bars, experimental group; open bars, control group.

as high as these make binocular single vision improbable, if not impossible.[23] That there were relatively more cases of the larger amounts of aniseikonia in the experimental than in the control group is further demonstrated by a comparison of the average amounts presented by each group, which were 2.08 percent in the case of the former and 1.49 percent in the case of the latter. The difference between these averages, which are based on only those subjects in each group who had 1.00 percent or more of aniseikonia, is 3.10 times the standard error of the difference.

[23] Ames, A., Jr., Gliddon, G. H., and Ogle, K. N., The importance of the relative size and shape of ocular images in vision, *Ann. Distinguished Service Found. of Optometry*, 1932, 1, 55–56.

Comparison of Experimental and Control Groups for Incidence of Aniseikonia at Near Vision and at Distant Vision.—It has been observed clinically that measurements of aniseikonia secured at one test distance may, or may not, agree with those obtained at the other. For example, some individuals may have aniseikonia at 20 feet, but not at reading distance; others may have it at reading distance, but not at 20 feet; and still others may have it at both distances, but in different amounts. A few cases have even been observed in which the image of one eye was smaller relative to the size of that of the other at one test distance, but larger at the other test distance. Theoretically, when the aniseikonia is present only at 20 feet, it should produce much difficulty at the movies, but not in reading. Similarly, when the aniseikonia is present only at the near point, it should produce disturbances when the eyes are used in reading and in other close work, but not when the eyes are used for distant vision. If the above theoretical considerations are correct, it would seem that measurements of aniseikonia secured at the reading distance should differentiate between groups of good and poor readers more clearly than would those obtained at 20 feet. That is exactly the case, as the analysis presented below will show.

The frequency with which various amounts of aniseikonia occurred at each of the two test distances is shown for 68 subjects in both the experimental and control groups on Table 3. That a larger number of cases was not used in this comparison is explained by the fact that aniseikonic findings at both test distances were available on but 68 of the original 100 pairs of subjects. The number of cases is, however, sufficient for the present purpose. An inspection of Table 3 will reveal that there were relatively more cases with aniseikonia of 1.00 percent or more in the experimental than in the control group at both test distances. Expressed in terms of percentages, the figures were 56 and 40 for the experimental group and 22 and 24 for the control group at the near and far points, respectively. The ratio of the difference between these percentages to the standard error of the difference was

4.36 at reading distance and 2.03 at 20 feet. If we accept a ratio of 3.00 or greater as indicating a difference which is larger than could occur under random sampling, it will be observed that only the difference at reading distance was statistically significant. As we above indicated, this is precisely the way in which the results should turn out, if an aniseikonia at the near point, only, operates to produce

TABLE 3

DISTRIBUTION OF EXPERIMENTAL AND CONTROL CASES ACCORDING TO AMOUNT OF ANISEIKONIA AT NEAR VISION AND AT DISTANT VISION

Aniseikonia in Percent	Frequency			
	Experimental		Control	
	Near	Distant	Near	Distant
0.00.....	10	13	12	8
0.25.....	5	7	18	17
0.50.....	9 44%	11 60%	12 78%	16 76%
0.75.....	6	10	11	11
1.00.....	5	6	5	6
1.25.....	4	7	4	2
1.50.....	6	4	1	2
1.75.....	3	1	1	2
2.00.....	4	2	2	2
2.25.....	4	0	1	1
2.50.....	2	1	0	1
2.75.....	0 56%	0 40%	1 22%	0 24%
3.00.....	6	3	0	0
3.25.....	0	0	0	0
3.50.....	0	2	0	0
3.75.....	1	0	0	0
4.00.....	1	0	0	0
Over 4.00..	2	1	0	0
Totals...	68	68	68	68

visual discomfort during reading. That the difference at 20 feet was also somewhat larger than can be accounted for on the basis of chance alone is immaterial. The critical finding is that the measurements of aniseikonia secured at reading distance differentiate our groups more significantly than do those obtained at distant vision. That the results from the test at 20 feet differentiate the groups at all is to be explained as an incident of the fact that there is a tendency for the results at the two test distances to agree, despite the

exceptions which are occasionally observed clinically. Some indication of the amount of agreement which might be expected in non-clinical cases is given by a comparison of the distributions for the control group on Table 3. It will be noted that the number of cases with aniseikonia of 1.00 percent or more was approximately the same at near and at distant vision, the figures in the two cases being 15 and 16, respectively. Twelve of the individuals in each of these groups were the same. Four cases had 1.00 percent or more of aniseikonia at distant vision, but not at near; three had this amount at near, but not at distant vision. Of these seven cases, the one which showed the greatest discrepancy was that of a twelve year old boy who had 1.75 percent of aniseikonia at the near point, and only 0.50 percent at the far point. The average amount of change from one test to the other for the 68 subjects was slightly less than 0.25 percent. A few individuals varied somewhat more than this, of course, but that the tendency on the part of the majority of cases was toward agreement of results is indicated by the fact that only 14 of the 68 subjects varied from one test to the other by amounts greater than could be accounted for in terms of errors of measurement. It might be remarked, in this connection, that the change in 10 of these cases was in the direction of a higher amount of aniseikonia at the far point. We should finally state that the correlation between the results at the two test distances for these 68 control cases was rather high, being $.79 \pm .03$.

In the light of the agreement which obtains between aniseikonic measurements at the near and far points in subjects more or less randomly selected, it seems even more significant that the results in the case of the experimental group varied from one test distance to the other as much as they did. The direction of this variation is shown graphically on Fig. 4. As on Fig. 3 the amounts of aniseikonia expressed in percent are plotted along the abscissa and the frequency of occurrence of these amounts along the ordinate. At nine of the thirteen points at which frequencies were plotted for amounts of aniseikonia in excess of 0.75 percent, the number of cases

presented was greater for near than for distant vision. The number of cases presented for distant vision was greater at only three points, namely, at 1.00, 1.25, and 3.50 percent. One case occurred for each distance at 5.00 percent. There were 38 cases, in all, with amounts of aniseikonia in excess of 0.75 percent for near vision as compared to 27 with these amounts for distant vision (*cf.* Table 3). These figures represent, respectively, 56 and 40 percent of the total number of cases in each distribution. When an adjustment is made for the correlation (.63 ± .05), which exists between the measurements at the two test distances, the difference between

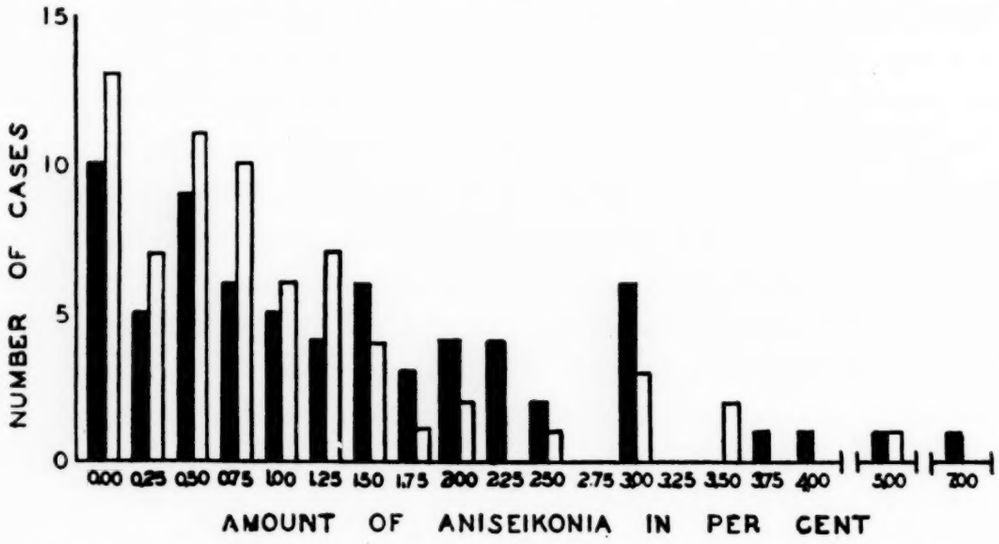


FIG. 4. Comparison of aniseikonic findings at near and at distant vision for experimental group. Solid bars, near vision; open bars, distant vision.

these percentages is 3.14 times the standard error of the difference. If this ratio be converted into terms of probability, the chances are 99.9+ in 100 that the observed difference is greater than can be attributed to random fluctuations.

The finding that disability in reading is rather more directly related to aniseikonia at the near than at the far point is further emphasized by an analysis of results obtained in individual cases. There were 39 cases, in all, with significant amounts of aniseikonia at one, or both, of the test distances. Of these, 12 had significant amounts at near, but not at distant vision; 15 had significant amounts at both distances, but the amounts presented at near were larger; and in 10 cases the amounts presented at the two distances

were the same. Of the remaining two cases, one had a significant amount only at the far point; the other had significant amounts at both distances, but the amount presented at the far point was larger. Thus, 27, or 71 percent, of the 38 subjects (we have excluded from this analysis the one subject who had a significant amount of aniseikonia only at distant vision), whose difficulties in reading might be attributable to aniseikonia, either had it only at near vision, or, if they had it at both distances, the amounts presented at near vision were larger. The average difference between the results at the two test distances for these 27 cases was 1.34 percent, the standard deviation of the distribution of variances being 0.63 percent.

Discussion

It need hardly be argued that our data have shown a significant relationship between disability in reading and aniseikonia, especially when measurements of the latter are obtained at reading distance. This fact is made obvious by even the most crude analysis of our data. How important aniseikonia is as a factor in the causation of reading disability is difficult to estimate, however. Certainly, it is not the only cause, for not all poor readers have it—nor is it seriously expected that they should. It is highly probable, moreover, that in most, if not all, of the poor readers who do have clinically significant amounts of aniseikonia other factors operate to a greater or lesser extent to complicate the problems presented. Severe disability in reading almost always occurs as a consequence of the accumulative influence of many factors, not on the basis of the effect of one alone. That being so, it is extremely difficult to isolate a single factor, such as aniseikonia, and to evaluate its importance independently of other factors which might be operating in a given case. However, if we accept—and we must—the condition of aniseikonia as one of the many factors which may contribute to failure in reading, it would appear, according to our data at least,[24] that it is a factor to be taken into account in ap-

[24] There is no reason for believing that our sampling of reading disability cases is atypical, insofar as the frequency with which aniseikonia occurred is concerned.

proximately 50 percent of extreme cases of reading disability. Not all of these cases will be similarly handicapped, even when the amounts of aniseikonia presented are the same or nearly the same, because some will have a greater tolerance for the defect than others. And in any event, aniseikonia is not an insurmountable handicap, as is evidenced by the fact that 23 percent of the cases in the control group had clinically significant amounts of it. We know that many of these cases were reading as well as their mental ages indicated that they should. One must not underestimate the importance of aniseikonia on the basis of these remarks, however, for the same statements can be made in connection with any of the structural factors which condition progress in reading, such as errors of refraction, muscular imbalance, deviations in ocular and manual lateral dominance, and so forth. If anything, our findings indicate that aniseikonia is a more frequently encountered obstacle than many of these factors. Other eye defects, for example, have not been found in recent studies to differentiate groups of good and poor readers as significantly as our study shows aniseikonia to do. That this should be the case is not surprising, for aniseikonia is an anomaly of the binocular visual apparatus, the effect of which becomes prominent only when the two eyes are engaged in a visual act such as reading. Refractive errors (near- and far-sightedness, and astigmatism), on the other hand, exist monocularly and affect the coördination of the eyes only indirectly, if at all. In fact, a slight near-sightedness may make it easier for a person to read because he is already focussed for near work. Far-sightedness is the more grave defect because it requires an exercise of the accommodation

Although the reputation of a clinic determines to some extent the type of case coming to it, it cannot be argued that a selective influence operated in the present instance to give us a disproportionate number of aniseikonic cases, notwithstanding the fact that it might have been more or less generally known in the region from which most of our cases were drawn that we were equipped in the laboratory of the Clinic to make measurements of aniseikonia and that we believed the condition to be of some importance in the diagnosis of reading disability. It is quite impossible for a naive person to tell, or even to guess, in advance of being measured for it, that he has aniseikonia, nor is an outsider in any better position to judge that another person has aniseikonia. Our clinical cases could not, therefore, have been selected.

for near work which may produce ocular fatigue. The eye defects which most seriously affect the ability to read are those which result in fusion inadequacies. As we indicated in our introductory statement, that is precisely the way in which aniseikonia is conceived to interfere with vision.

Summary and Conclusions

Analysis of the records of examinations for aniseikonia of 100 cases of severe disability in reading and of 100 control cases has shown the following:

1. Fifty-one percent of the experimental cases and 23 percent of the control cases had clinically significant amounts of aniseikonia. The difference between these percentages is 4.31 times its standard error.

2. Of the cases in both groups with significant amounts of aniseikonia, those in the experimental group had amounts ranging from 1.00 to 7.00 percent, with an average of 2.08 percent; whereas those in the control group had amounts ranging from 1.00 to 2.75 percent, with an average of 1.49 percent. The difference between these averages is 3.10 times the standard error of the difference.

3. Of the 68 paired cases in each group on whom aniseikonic measurements were available at both test distances, 56 percent of the experimental cases and 22 percent of the control cases had significant amounts of aniseikonia at reading distance, while at 20 feet 40 percent of the former and 24 percent of the latter had these amounts. The ratio of the difference between these percentages to the standard error of the difference is 4.36 at the near point and 2.03 at the far point.

The greater frequency with which the larger amounts of aniseikonia occurred among the experimental than control cases and the fact that the measurements secured at reading distance differentiated the groups more significantly than did those obtained at 20 feet justify the conclusion that aniseikonia is one of the many factors that may contribute to the causation and persistence of disability in reading.

(Manuscript received June 27, 1938)

STUDIES IN THE TRANSPOSITION OF LEARNING BY CHILDREN[1]

I. Relative vs Absolute Response as a Function of Amount of Training

BY T. A. JACKSON, E. STONEX, E. LANE, AND K. DOMINGUEZ

Columbia University

I

In the past three or four decades, a score or more of studies have been made which bear upon the general problem of absolute and relative choice in the transposition experiment. One of the earliest of these was that of Kinnaman (**11**) who found the relative reaction to predominate. Subsequent experimental work showing essentially the same result was done by Casteel (**3**), Breed (**2**), Washburn and Abbott (**21**), Bingham (**1**), Coburn (**4**), Johnson (**9**), and Gulliksen (**5**). Kohler's (**12**) work with hens marked the beginning of a more vital interest in the problem. It was the Gestalt School that emphasized the broad theoretical implications of the phenomenon. The interpretation they gave was that the relative response was 'more stable.' It was implied that the absolute choice was an exception, was temporary, and occurred only under special experimental conditions. One of the conditions that would induce it, according to Kohler, was a short interval between the training and the critical trials.

A number of more recent studies agree with the earlier work in finding the relative response more prevalent than the absolute. Jones and Dunn (**10**) made a significant contribution in finding a positive correlation between the tendency consistently to make the relative choice and the speed of original learning. Takei (**18**) and Taylor (**17**) found, con-

[1] The senior author is responsible for the form of the present report. Under his direction the junior authors collected the data presented in connection with their work on the Master's essay which are on file in the Columbia University Library.

trary to the usual result, that absolute choice occurs more or less consistently if the stimuli are sufficiently different. They used black, gray, and white stimuli instead of three grays. The work of Warden and his collaborators (**19**, **20**), although not strictly comparable to the usual transposition experiment, due to their use of generalized training, tends to favor the absolute choice. They suggest that the type of response made is a function of the method of testing employed.

Gundlack and Herrington (**6**) have emphasized the importance of analyzing the conditions under which the relative or the absolute choice occurs. They stress particularly the following factors; first, the size of the time interval between the two presentations of the stimuli (in successive presentation), and second, the difference between the stimuli. Klüver (**13**) also points out the great need of experimental analysis of the problem. The experimental work of the latter is by far the most extensive and systematic of any to date on this general problem. His work is a careful analysis of the 'equivalence' of stimuli in the discrimination learning situation. In other words he attempted to determine how much stimuli could be varied in size, shape, color, weight, etc., and still obtain the same relative choice after the changes were made. Wide ranges of so-called 'equivalence' were found in three different sense modalities.

II

The aim of the present experiments is to study systematically the conditions of relative and absolute choice when the stimuli have been transposed. Particular attention will be given to the factor of degree of practice on the original pair of stimuli. If, in discrimination learning, the stimuli exert a generalized effect after some practice, and with more and more practice a less and less generalized effect, then it should be possible to predict on the basis of these generalizations the amount of practice that would yield a relative response upon transposition of the stimuli and also the amount that would make for absolute choice. Since we do not know the shape of the curve that would represent the generalization at any

stage of practice, nor the rate of change in the curve with increased differentiation, we shall attempt only to make a general prediction. In discrimination learning with stimuli that are easy to discriminate, following a relatively small amount of practice beyond fixation,[2] a relative choice will be made, and after a relatively great amount of practice the absolute choice will occur. The basis for this general line of reasoning comes mainly from the following sources. First of all Wiltbank (**22**) and Jackson (**8**) have shown that, in maze learning, differential transfer effects follow different degrees of practice on the first maze. The latter suggests an interpretation of his results in terms of the concepts of generalization and differentiation as used by Pavlov (**14**) and Razran (**15**). In a study by Spence (**16**) an excellent analysis of transfer effects is made by use of hypothetical generalization curves. After chimpanzees were trained to select the larger of two visual stimuli, which were 160 and 100 millimeters square, respectively, they were given critical trials on the following pairs of stimuli: 256 and 160, 320 and 200, 409 and 256. Each of these pairs has the same interstimulus relation as the training pair. The average percentages of relative choice were 88.3, 80.0, and 78.3 respectively. These results are consistent with the expectations demanded by the generalization curves. Of particular significance for our problem is the recent work of Hovland (**7**) who found that differentiation takes place without differential reënforcement. His study clarifies our understanding of one of the basic mechanisms involved in transfer where practice is a variable.

The Gestalt psychologists will doubtlessly criticize the general line of attack used by Spence and by the present writers for making the analytical error, that is, for breaking up the response made to a total situation into two elements, one corresponding to each stimulus. In constructing his hypothetical generalization curves, Spence made use of

[2] By the term 'fixation' is meant that point in learning a discrimination habit beyond which no errors are made (at least practically). The term 'post-fixation' will then designate the practice which takes place during and following the norm. It is hoped that the latter term will be an acceptable substitute for the unacceptable 'overlearning,' which has been vigorously criticized in certain quarters.

principles that were developed in work on the conditioned response which appears to be quite different from the discrimination learning situation. However, Spence's approach has the virtue of providing a basis for testing specific hypotheses quantitatively. In spite of the 'apparent' difference between conditioning and discrimination learning, the principles of generalization and differentiation probably apply in the latter. We shall make that assumption until evidence proves it to be unsound.

III

Apparatus

The apparatus presented two visual stimuli to the subject simultaneously. The stimuli were attached to the ends of square rods which were placed vertically twelve inches apart on top of the box. With such an arrangement there was no immediate background the subject could use as an aid in making absolute judgments of the stimuli. Comparisons could be made only between the two stimuli. The two rods were hinged on the box so that they could be pushed back about an inch or two. A subject indicated his choice by pushing back one or the other of the stimulus patches. If the correct one was chosen, a bell tinkled and the reward box in front of the rod opened; if the wrong choice was made nothing happened, as the stimulus was pushed back. The apparatus was set for right or left response by adjustments on the back of the box out of sight of the child. The screen, which could be erected by manipulation of the lever on the board back of the box, was always raised between trials. In the first few experiments the food boxes were not used and so were kept locked all through these trials. Certain modifications were made in the apparatus for the last of the experiments reported. These changes will be described later.

The stimuli used were made of squares of wood, their areas being 1600, 4900, and 15,000 sq. mm. respectively. In introducing the task to the child, certain preliminary training was given with two other stimuli, namely a circle (5 centimeters in diameter) and an equilateral triangle (measuring 5 centimeters on a side). In experiments where other stimuli were used, mention will be made of them in connection with the corresponding results.

Subjects

In the four experiments reported, forty-five different children wereuse d. They ranged in age from three to six years and were obtained in various kindergartens and day nurseries in New York City.[3] The reason for using young children was to avoid the possibility of transfer by verbal analysis. One might suppose that we should use children that have not learned to talk; however, our general observations seem to indicate that although the subjects may readily understand the concept 'bigger than,'

[3] We wish to take this opportunity to express our very sincere appreciation to the superintendents and staffs of the following day nurseries for their coöperation: The Bethany Day Nursery, The Washington Heights Day Nursery, The Manhattanville Day Nursery, and the Kindergarten of the Washington Heights Y. M. and Y. W. H. A

they do not transfer by such verbal analysis on critical trials. Language did not appear to be a factor in making a choice between the transposed stimuli among our subjects.

IV

Experiment One

The particular problem in this experiment was to determine the effect of different amounts of practice of a visual discrimination habit upon the type of response made in critical trials with the transposed stimuli. The apparatus and stimuli used was as described above. The two smaller stimuli were used in training and transposition was made to the larger pair. In both pairs the larger stimulus is 75 percent greater than the smaller, in linear dimensions, and ratios of areas was about one to three. The subjects were twenty-three kindergarten and nursery school children, ranging in age from three to six years. There were twelve boys and eleven girls.

The task was presented to the subjects as a game. The child was told to try to ring the bell; and if the right stick were pushed back, the bell would ring. This sort of incentive was thought to be sufficient to interest the child. Before the regular training two other stimuli, a circle and a triangle, were used to introduce the task to the child. In one school if the subject did not learn the preliminary discrimination within two days, (about twenty trials) that child was not used. However, in the second school they were changed to the regular training after two days regardless of whether they had learned the first. This procedure eliminated eight out of the original twenty-three subjects. Irregular attendance made it necessary to drop three others, leaving twelve upon whom data were obtained.

Reaction time was taken with a stop watch. The time interval in question began with the lowering of the screen and ended when one of the two stimuli was reacted to and pushed back far enough to ring the bell or the equivalent distance. After the regular training only one critical trial was given, and whichever choice was made the child received the reward. A second critical trial would obviously be influenced by the reward given on the first as has been pointed out by Warden (**20**). In these experiments training was to the larger stimulus and transposition was upward. In both training and transposition trials the stimuli were presented in a systematic haphazard order with respect to right or left position of the positive stimulus. The norm used throughout the four experiments was four correct in five trials.

Results

Out of the fifteen subjects, only two made an absolute response on the critical trial. One of these followed a small amount of practice beyond fixation, and the other followed a relatively large amount. It is impossible on the basis of these data to make any conclusions regarding the relation between amount of practice and the kind of response made upon transposed stimuli. However, there is a suggestion in the case of one subject (Mi) who made the relative response after twenty trials and an absolute response after fifty trials. This may be a chance error, but since it is in line with what

might reasonably be expected it becomes worthy of note. The differentiation produced by the additional practice may have been just sufficient to change the choice upon the critical trial from relative to absolute. However, if this were the case the same interpretation will not cover the other instance of an absolute choice, for it was made following only ten trials of practice beyond fixation. It is not necessary that the two

TABLE 1

SHOWING THE TYPE OF RESPONSE MADE UPON TRANSPOSITION, AND REACTION TIME IN BOTH THE PRACTICE AND CRITICAL TRIALS

Subject	Age in Years	No. Trials Practice		Response on Crit. Trial	Reaction Time in Seconds	
		Total	Post-Fixation		Practice	Critical
Gl.....	5	13	5	Relative	3.6	3.0
Al.....	3	17	10	Relative	1.9	5.0
Ra.....	6	24	10	Relative	1.0	1.1
Jl.....	5½	20	10	Absolute	1.2	2.7
Ol.....	4½	23	20	Relative	1.1	2.0
Le.....	4	53	20	Relative	1.6	2.0
Jn.....	5½	23	20	Relative	1.2	3.0
Mi.....	6	26	20	Relative	1.7	3.5
Ge.....	6	20	20	Relative	1.5	1.5
Ed.....	5½	55	40	Relative	1.8	2.0
Ma.....	5½	48	45	Relative	1.3	1.5
Mi*.....	6	56	50	Absolute	1.4	2.1
Ra*.....	6	74	60	Relative	1.3	2.0
Mr.....	5	80	80	Relative	1.5	2.0
Ma*.....	5½	98	95	Relative	1.3	2.0
				Average	1.56	2.36

* These three subjects were used twice.

instances be subsumed under the same explanation. It may well have been that in the latter case the habit of choosing the larger stimulus had not been thoroughly fixed or stable at ten post-fixation trials. The shock of transposition was too great for such a weak habit, and the subject regressed to trial and error behavior. The foregoing interpretation may seem quite gratuitous but we shall be able to support it with data reported later in this study.

The reaction time data show that the critical trials required more time than the training trials. Such was the

result in fourteen of the fifteen cases. The averages are 1.56 and 2.36 seconds respectively. The critical trials required 51 percent more time on the average. The difference is a relatively great one and it indicates, we believe, that the transposed stimuli are not entirely 'equivalent.' When the transposed stimuli are presented the subject pauses momentarily for some sort of readjustment, and only after that does he make a choice. Only if one ignore the temporal aspects of the responses can they be regarded as equivalent. An appearance of equivalence is manifest because the situation is such as to force a response into one or the other of only two patterns.

Summary of Experiment One

Fifteen children were trained to react to the larger of two square stimulus patches. The incentive for such reaction was hearing a bell ring and otherwise playing a game. After different amounts of practice of this discrimination habit, the stimuli were transposed from a pair having areas 1600 and 4900 sq. mm., to a pair having areas of 4900 and 15,000 sq. mm. respectively. Only one transposition trial was given. Of the fifteen cases thirteen made the relative choice, that is, reacted to the larger stimulus of the second pair. The reaction time in the critical trials was 51 percent greater than in the training trials.

V

Experiment Two

This experiment is essentially a repetition of the first. The same apparatus and stimuli were used. However, several refinements were made in procedure. The experimental work in this and the two succeeding experiments was carried on in a small room apart from the other children at play. The work was expedited by having two experimenters present during all trials. One operated the apparatus and the other recorded the data and took the reaction time. This change prevented losing rapport with the child, and the trials were sufficiently frequent that the subject's attention wandered less. The second change of procedure was that all post-fixation trials were given on one day, and the critical trial followed immediately. Usually the whole test took place on one day. The third change was to make the smaller of the pair of stimuli positive. Accordingly, the training squares were 4900 sq. mm. and 15,000 sq. mm. and the transposition squares were 1600 sq. mm. and 4900 sq. mm. No preliminary trials were given with the circle and the triangle stimuli.

Nine children were used from one of the same schools visited in the previous experiment. They were divided into three groups of three children each. They were given respectively, five, ten, and twenty trials beyond fixation.

Results

These results, see Tables 2 and 3, show a distinct and consistent relationship between amount of post-fixation

TABLE 2

Showing the Type of Response Made upon Transposition, and the Reaction Time in Both Practice and Critical Trials

Subject	No. Trials Practice		Response on Crit. Trials	Reaction Time in Seconds	
	Total	Post-Fixation		Practice	Critical
A. L.	14	5	Relative	4.6	7.0
G. E.	5	5	Relative	2.8	4.1
R. O.	9	5	Relative	4.3	3.0
C. H.	24	10	Absolute	1.8	2.4
R. A.	15	10	Relative	5.9	4.0
R. X.	13	10	Relative	3.5	5.0
A. F.	22	20	Relative	1.9	3.9
R. L.	47	20	Absolute	2.7	3.9
T. A.	22	20	Absolute	1.3	1.0

TABLE 3

Summary Showing the Type of Response Made upon Transposition, and the Reaction Time in Both Practice and Critical Trials

Group *	No. Trials Practice		Response on Critical Trials		Reaction Time in Seconds	
	Ave. Total	Post-Fixation	No. Rel.	No. Abs.	Ave. Practice	Ave. Critical
I.	9.3	5	3	0	3.9	4.7
II.	17.3	10	2	1	3.7	3.8
III.	30.3	20	1	2	2.0	2.9
					Ave. (3.2)	(3.8)

* Three subjects in each group.

practice and the type of choice made in the transposition trial. The greater the amount of practice, the more absolute responses were made and the fewer relative.

The differences between these results and those of the previous study are such as to indicate some explanation. The

work in the present experiment was more carefully carried out. This was possible because two experimenters were present and the child's attention was retained throughout the whole procedure. In the earlier work some delays were unavoidable.

Another difference between the two experiments is in the distribution of practice. Spence and others have suggested that transfer following massed practice would not be the same as following distributed practice. In our first experiment the practice was distributed over four or five days, but in this one it was all done on one day. Although this is undoubtedly an important factor in transfer, the results of our two experiments are not such that we can throw any light on the question either way.

Perhaps the most important difference between these two studies lies in the matter of the direction of transposition. Since there appears to be a preference for the larger stimulus among children of the ages employed, the absolute choice would occur more frequently in transposition downward and the relative choice in transposition upward. This would account for the two absolute choices following only twenty post-fixation trials. This bias favoring the absolute choice in transposition downward does not in any way invalidate the general conclusion that the absolute choice is made after more practice than the relative. The point of transition from relative to absolute occurs earlier in transposition downward than in the opposite direction.

The reaction time results add very little to the findings of the earlier experiment. The average reaction time is greater for the critical trials than for training trials. They are 3.8 and 3.2 respectively.

Summary of Experiment Two

Nine children, ranging in age from four and a half to six years, were trained to react positively to the smaller of two square stimulus patches (4900 sq. mm. and 15,000 sq. mm.). They were divided into three groups of three children each and given five, ten, and twenty post-fixation trials, respectively. One trial on transposed stimuli (1600 sq. mm. and

4900 sq. mm.) was given each subject. All the first group made relative choices; of the second, two made relative choices and one an absolute; of the third group only one made the relative choice. The reaction time data again showed that more time is required on the critical trials than on the training trials.

VI

Experiment Three

The general problem in this experiment is the same as in the two previous ones, but weights were used as stimuli. They were metal cylinders eight and a half cm. high and four cm. in diameter, weighing 30, 120, and 480 grams respectively. The apparatus used in the two previous experiments was employed in this one but without any visual stimuli attached to the rods. Two of the weights were set on the box, one near each of the two square rods. The subject was instructed to lift one weight in each hand simultaneously, then to push the rod next to the one he thought was the 'right' one to make the bell ring.

Nine children were used from the same group as Experiment II, three having acted as subjects in both. The general procedure was the same as in II, with one exception. Training was given on the smaller pair of stimuli, and transposition was made to the larger pair.

Results

The results of this experiment (Tables 4 and 5) support the general thesis set forth earlier in the report and are in general agreement with the results of the second experiment.

TABLE 4

Showing the Type of Response Made upon Transposition, and the Reaction Time in Both Practice and Critical Trials

Subject	No. Trials Practice		Response on Crit. Trials	Reaction Time in Seconds	
	Total	Post-Fixation		Practice	Critical
I. R.	15	5	Relative	*—	—
R. L.	8	5	Relative	6.8	7.0
T. A.	8	5	Relative	3.8	6.0
B. E.	37	10	Relative	—	—
J. U.	18	10	Relative	—	—
R. O.	10	10	Relative	2.9	5.0
E. D.	20	20	Relative	3.4	8.0
M. R.	24	20	Relative	2.8	3.5
M. Y.	20	20	Absolute	5.0	8.5

* Reaction time measures were inadvertently omitted in three subjects.

TABLE 5

SUMMARY SHOWING THE TYPE OF RESPONSE MADE UPON TRANSPOSITION, AND THE REACTION TIME FOR BOTH PRACTICE AND CRITICAL TRIALS

Group	No. Practice Trials		Response on Crit. Trials		Ave. Reaction Time in seconds	
	Ave. Total	Post-Fixation	No. Rel.	No. Abs.	Practice Ave.	Critical Ave.
I.....	10.3	5	3	0	5.3	6.5
II.....	21.7	10	3	0	2.9	5.0
III.....	21.7	20	2	1	3.7	6.7

Although there was only one case of an absolute choice in the critical trials, it is significant that it occurred in the group where it was most expected. The data presented suggest that the transition point from relative to absolute choice in the case of these stimuli occurs after somewhat more practice than with the visual size relata when transposition is downward.

The reaction time data are also generally consistent with those of Experiment II. Although it may be hazardous to make too much of such scant results, it is at least suggestive that the relative difference between the two reaction time measures (Practice Ave. vs. Critical) is greater in the case of more post-fixation practice. The percent increase for the criticals trials as compared with the practice is 23, 72, and 81 for groups I, II, and III respectively. This relationship is due, probably, to the greater conflict between the two alternative responses under the condition of greater practice.

SUMMARY OF EXPERIMENT THREE

Nine children were trained to choose the heavier of two weights (30 vs. 120 grams). They were divided into three equal groups and given various amounts of post-fixation practice of the discrimination habit; five, ten, and twenty trials were given to groups I, II, and III respectively. Each subject was tested upon transposed stimuli (120 vs. 480 grams) immediately following the last training trial. All subjects that had five or ten trials of post-fixation practice made the

relative choice. One of the three with twenty trials made an absolute choice in the transposition test. There was some evidence of greater relative delay in making choice between the transposed stimuli in the groups with longer practice.

VII

Experiment Four

Two variables were used in this experiment; one, the effect of degree of practice as previously, and the other, the effect of the degree of difference between the stimuli. Whereas in the foregoing studies the ratio between the areas of the stimuli was about $\frac{1}{3}$ (1600 to 4900), in this study ratios $\frac{1}{4}$, $\frac{1}{2}$, and $\frac{3}{4}$ were employed. The first of these was very easily learned by our subjects, but the last ($\frac{3}{4}$) was so difficult that work with it was finally discontinued and no data are presented for it.

The apparatus used was a modification of the one used in the foregoing experiments. Two features of the original model had to be given up in order to make the apparatus more flexible, that is, more easily and speedily used. The screen was greatly simplified. A cross arm attached perpendicularly to the middle of the screen was hinged to the middle of the background so that the screen would pivot over the top of the background. This type of screen was found to be less disturbing to the subjects. The other modification was the introduction of a black background for the stimuli. This was necessary because the variety of stimuli desired could not easily be attached to the posts of the other apparatus. The stimuli were made of paper, then pasted on heavy cardboard, and glued to holders made of small strips of brass. The stimuli were held in place by inserting the holders into small grooves through the background. The holder just touched the posts; and when the subject pushed on a stimulus, the post was moved back just as in the earlier model of the apparatus. The reward boxes were not changed. The stimuli were octagonal in shape instead of square as previously.

The subjects used in this experiment were from one of the schools mentioned in earlier sections of the report. Twelve children were used. They ranged in age from twenty-nine to seventy months. The two main groups were equated for age.

The general procedure was not changed except as follows. The incentive conditions were modified by the introduction of a material reward. The child received a raisin in each trial in which he made a correct choice. These were presented to him by means of the reward boxes mentioned earlier. On the first day or two of experimentation the raisins were eaten immediately but after that they were saved. The children apparently preferred to keep them so as to show their fellow pupils as well as to share the earnings. A short preliminary test was given each subject; this consisted of ten trials with stimuli as different as possible from those of the main experiment. A triangle and a square were used for this purpose. It was not possible to give all training trials on one day, but they were distributed as little as possible. Training rarely covered more than two days for the subjects given thirty post-fixation trials. In all cases at least twenty of the training trials occurred just preceding transposition. The subjects given one hundred trials beyond fixation required about five days of work altogether. It was decided to give three trials with the transposed stimuli instead of one as previously; however, the results will be kept in separate categories.

Results

Using the 500, 2000 and 8000 sq. mm. stimuli (ratio $\frac{1}{4}$) and under the condition of thirty trials after fixation, all four subjects made relative choice through three critical trials. Following one hundred trials after fixation, one of the two subjects made three absolute responses, and the other three relative choices. In all these cases selection seemed to be

TABLE 6

Showing the Type of Response Made from Transposition and the Reaction Time in Both Practice and Critical Trials

Subject (Age in Months)	Ratio of Stimuli	No. Practice Trials		Response on Critical Trials			Reaction Time in Seconds			
		Total	Post-Fixation	1st	2nd	3rd	Practice *	Critical Trials		
								1st	2nd	3rd
Ai (32)	$\frac{1}{4}$	48	30	Rel.	Rel.	Rel.	1.7	2.0	1.5	1.5
Ke (45)	$\frac{1}{4}$	39	30	Rel.	Rel.	Rel.	2.2	3.0	2.5	2.5
Ro (53)	$\frac{1}{4}$	34	30	Rel.	Rel.	Rel.	1.7	2.0	2.5	2.0
Ru (71)	$\frac{1}{4}$	43	30	Rel.	Rel.	Rel.	2.3	3.5	3.0	2.0
He (39)	$\frac{1}{4}$	†	100	Abs.	Abs.	Abs.	3.5	5.5	4.5	5.0
Ro (66)	$\frac{1}{4}$	†	100	Rel.	Rel.	Rel.	1.3	3.0	2.5	1.5
Ro (29)	$\frac{1}{2}$	43	30	Abs.	Rel.	Rel. (R) ‡	1.9	2.0	2.0	1.5
Pa (44)	$\frac{1}{2}$	48	30	Rel.	Rel.	Abs. (RAA)	1.5	2.0	2.0	1.5
Ph (52)	$\frac{1}{2}$	48	30	Rel.	Rel.	Rel.	1.9	2.5	1.5	2.5
An (70)	$\frac{1}{2}$	42	30	Rel.	Rel.	Abs. (RR)	1.8	3.5	2.0	2.5
Ju (41)	$\frac{1}{2}$	†	100	Rel.	Rel.	Rel.	2.0	2.5	2.5	2.5
St (70)	$\frac{1}{2}$	†	100	Rel.	Rel.	Rel.	1.5	2.5	2.5	1.5

* These Reaction Time averages were based on the last 10 practice trials.

† Record of total trials in these cases were inadvertently omitted.

‡ A few extra critical trials were given when the first three were not of the same type of response.

stable. In the results for the second set of stimuli (1000, 2000, and 4000 sq. mm.), under the condition of thirty post-fixation trials, three our of four subjects responded inconsistently or unstably on the critical trials. This is shown by the shifting back and forth from relative to absolute choice through a series of five or six trials. The results with a hundred trials show that both of two subjects made consistent responses in three critical trials, the choice being of the relative type. Selection under these conditions appears to be

stable and dependable, whereas it was not stable with less practice.

The results of this experiment make clear the relation between these variables. In case of the more easily discriminated stimuli (ratio $\frac{1}{4}$) the addition of seventy trials practice reduced the percentage of relative choice from one hundred to fifty. With the more difficult stimuli a stable relative choice was not made at the same degree of practice but did appear with more practice. In other words, to reach the same point of potential transfer effect, more practice is required with the less easily discriminated stimuli. It may

TABLE 7

SUMMARY SHOWING THE TYPE OF RESPONSE MADE UPON TRANSPOSITION AND THE REACTION TIME IN BOTH PRACTICE AND CRITICAL TRIALS

Ratio of Stimuli	No. Practice Trials		No. of Relative and Absolute Responses on Critical Trials						Ave. Reaction Time			
	Ave. Total	Post-Fixation	1st		2nd		3rd		Practice	Critical Trials		
			Rel.	Abs.	Rel.	Abs.	Rel.	Abs.		1st	2nd	3rd
$\frac{1}{4}$	41	30	4	0	4	0	4	0	2.0	2.6	2.4	2.0
$\frac{1}{4}$	†	100	1	1	1	1	1	1	2.4	4.1	3.5	3.2
$\frac{1}{2}$	45.3	30	3	1	4	0	2	2	1.8	2.5	1.8	2.0
$\frac{1}{2}$	†	100	2	0	2	0	2	0	1.7	2.5	2.5	2.0

be expected that still greater practice with the more difficult stimuli (ratio $\frac{1}{2}$) absolute choices would be made. However, as the relative difference between the stimuli is reduced, there is undoubtedly a point beyond which no amount of practice would produce a stable absolute choice upon transposition.

That there is a real difference between the results for the two degrees of practice and also between the results for the two sets of stimuli appears to be quite probable. However, in view of the small number of cases, no statistical analysis of the reliability of these results will be made.

The reaction time results are quite in line with the previous findings. There is a sharp increase from the last training trial to the first critical trial, but this increment gradually disappears in three or four critical trials. It is significant

that with the first set of stimuli (ratio $\frac{1}{4}$) there was a much greater increase in reaction time for one hundred trials than for thirty trials. The increments were 1.7 seconds and .6 of a second respectively. The greater increase in the former case indicates greater habit interference. The reaction time increment for the second set of stimuli at one hundred trials was .8 of a second which is about the same as resulted for thirty trials practice with the first set of stimuli. This we believe supports the foregoing interpretation.

Summary of Experiment Four

Twelve children varying in age from twenty-nine to seventy months were trained to choose the larger of two octagonal stimuli. Half of them were trained on stimuli which were 500 and 2000 sq. mm. respectively, then transposed to 2000 and 8000 sq. mm. stimuli; the other half were trained on a more difficult pair which were 1000 and 2000 sq. mm. respectively, then transposed to 2000 and 4000. Transposition followed thirty post-fixation trials in four subjects of each group, and one hundred post-fixation trials for two subjects in each group. A stable relative choice was made by all four subjects given thirty trials on the first set of stimuli, but with the second set only one of four subjects made the relative choice regularly through three critical trials.

Following one hundred trials practice stable choices were made with both sets of stimuli. Choice was relative for both subjects with the second set of stimuli, but one was relative and one absolute for the first set of stimuli. Reaction time was greater for critical trials than for practice trials.

VIII

Discussion

In the field of learning and transfer it is no longer sufficient merely to state one's findings with the reliabilities of the differences computed to the fourth decimal place. There is a sufficient body of fact to enable us to attempt some systematic formulation. Spence's valuable article mentioned earlier

has pointed the way and we shall apply the same type of analysis to our results.

Spence demonstrated that in discrimination learning the results obtained upon various degrees of transposition of the stimuli may be quite adequately interpreted in terms of irradiation effects about the positive and negative stimuli. These he showed graphically as generalization curves. The formula he used in constructing his curves is as follows: $E = 10 - 20\ D^2$ and $I = 6 - 20\ D^2$ where E equals excitation, I inhibition, and D the difference between the logarithms of the training and the testing stimuli.

In order to apply such an analysis to our data, it will be necessary to modify the formula so that these generalization curves will show less spread as the number of trials of practice increases. This can be done by putting some expression of the amount of practice in the formula. Certain constants in the formula were also changed. The formula giving what appears to be the best fit is as follows:

$$E = 10 - \frac{T(\log T)^2}{42.5} D^2,$$

$$I = 6 - \frac{T(\log T)^2}{42.5} D^2,$$

where T = number of post-fixation trials of practice.

An assumption involved in this formula, in addition to those stated by Spence, is that with practice there will develop a differentiation effect with respect to each stimulus. Hovland's work shows clearly that differentiation does occur in a conditioning set-up without a differential reenforcement. This particular assumption, then, is as well justified as is the whole procedure of applying conditioning principles to discrimination learning. The a priori argument for differentiation of the inhibition curve is only by analogy; however, our results give support to this assumption as will be shown.

The formula as applied to the conditions of our fourth experiment gives the curves shown in Figs. 1, 2, 3, and 4. The numbers between the two curves represent the net strength of the stimuli. This was obtained by algebraic

summation of the two curves. Referring to Fig. 1, we see that upon transposition to stimuli 2000 and 8000 the relative choice should occur because the value 5.66 for the larger stimulus is greater than 4.56 for the smaller. The question arises as to how much greater a value should be for prediction. The formula was adjusted so that, for these results, a difference greater than unity would indicate a stable choice to either relative or absolute.

We are now ready to check our results against the theoretical curves. Fig. 1 shows that we should get relative choice,

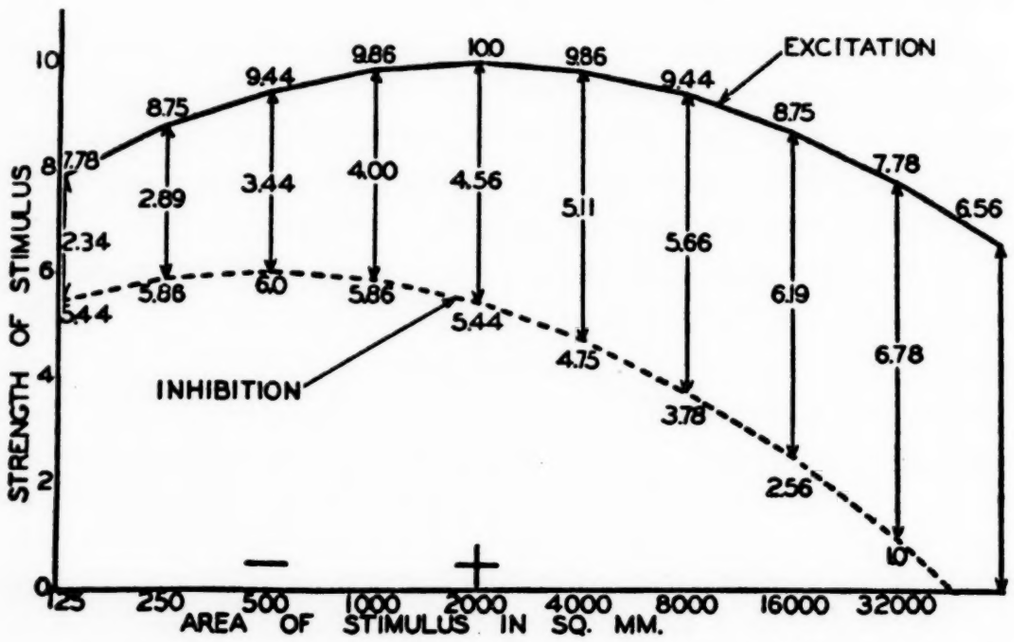


FIG. 1. Diagrammatic representation of the relationship between the hypothetical generalization curves, positive and negative, after 30 post-fixation trials of practice on the stimulus combination 500(−) and 2000(+).

the value being 1.10. Relative choice was obtained and all four subjects made that choice through three critical trials, indicating a certain stability in selection. Fig. 2, for one hundred trials, indicates an absolute choice but the value is less than one and we cannot be certain. One absolute and one relative choice was obtained. Both subjects were consistent in their choice through three critical trials. Fig. 3 indicates relative choice but the value, .28, is low. Three out of four subjects gave a relative choice on the first trial, but only one of the four subjects chose the larger stimulus con-

sistently through the three transposition trials. This of course is just as would be expected on the basis of the low value (.28). Finally, Fig. 4, yielding a value of 2.09, clearly indicates the relative choice, and stable relative choices were made by all subjects. Although these data fit the curves produced by the formula pretty well, we do not infer that it would predict results for other conditions unless appropriate changes were made in the constants. This sort of analysis seems to be good enough to justify further application to discrimination learning work.

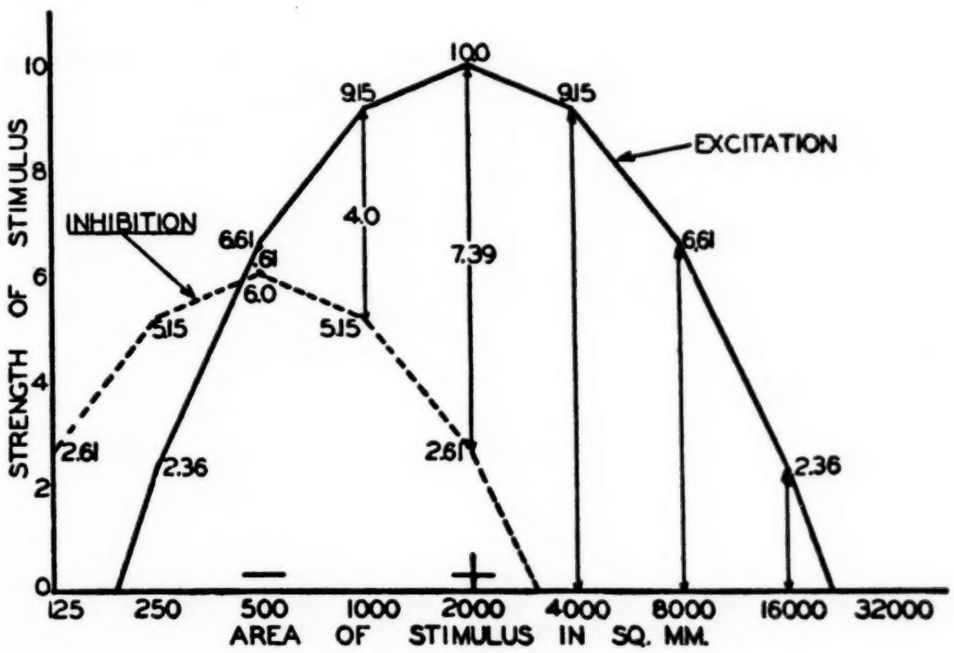


FIG. 2. Diagrammatic representation of the relationship between the hypothetical generalization curves, positive and negative, after 100 post-fixation trials of practice on the stimulus combination 500(−) and 2000(+).

The relation of the present study to previous work may now be considered. Jones and Dunn (**10**), in interpreting their results, have offered two explanations for their correlation between speed of learning and the tendency to make the relative response. First, they state that the relative response is characteristic of rapid learning, and second, that the relative response is an accompaniment of overlearning. According to the present interpretation neither of these are entirely correct and yet both are partly right. Our results show that the relative choice will be made following a certain degree of

practice. Rapid learners will reach that point more quickly than slow learners and obviously one might find a positive correlation such as they have reported, provided the learning is not carried too far beyond the optimum point for the relative choice. If practice had been considerably extended, they might have found a correlation between rapid learning and tendency to make the absolute response. With regard to their second explanation, it is already clear that the relative response is no more an accompaniment of overlearning (post-fixation practice) than is the absolute response; although the

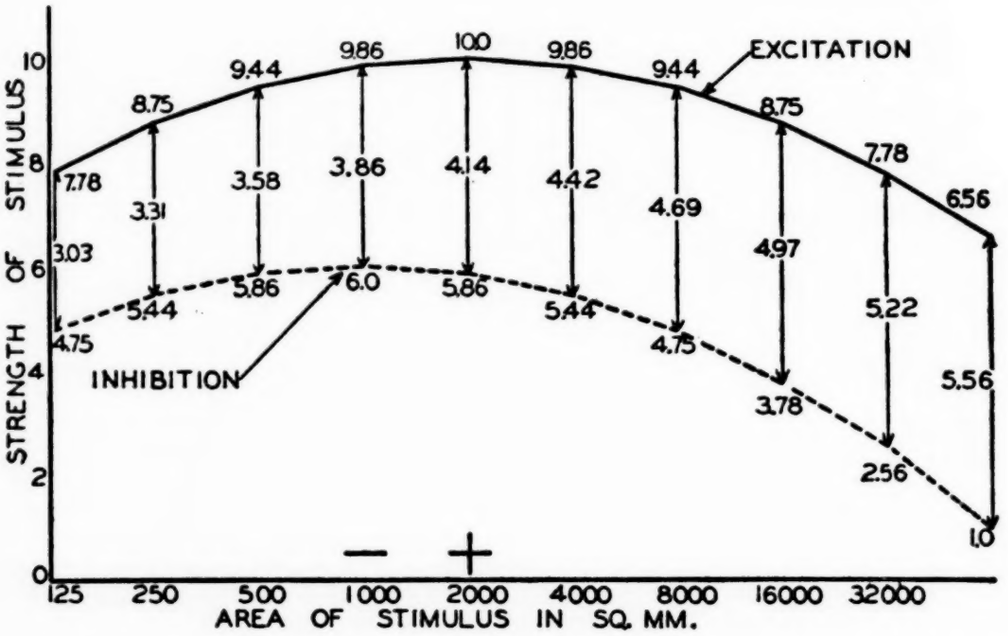


FIG. 3. Diagrammatic representation of the relationship between the hypothetical generalization curves, positive and negative, after 30 post-fixation trials of practice on the stimulus combination 1000(−) and 2000(+).

relative response will occur following smaller degrees of practice than the absolute. The degree of practice necessary to produce an absolute choice largely depends upon the amount of difference between the two stimuli. Our results show this to be the case, and Takei (**18**) and Taylor (**17**) obtained similar findings. They found that the absolute choice was made after only a moderate amount of practice, when stimuli that differ greatly were employed.

Klüver (**13**) has introduced the term 'equivalent stimuli' in his analysis of discrimination learning in the monkey. As

applied to our study this would mean that choosing the larger stimulus following transposition is equivalent to choosing the larger stimulus before transposition. In other words, the 4000 sq. mm. stimulus when paired with the 2000 sq. mm. stimulus is equivalent to the 2000 when paired with the 1000. (All of this is based upon the conditions of relative response.) Our results have consistently shown that more time is required to make choice following transposition than before. In the first critical trial subjects frequently remarked to the effect that "Well, this is something different! . . . Ah! . . . Which

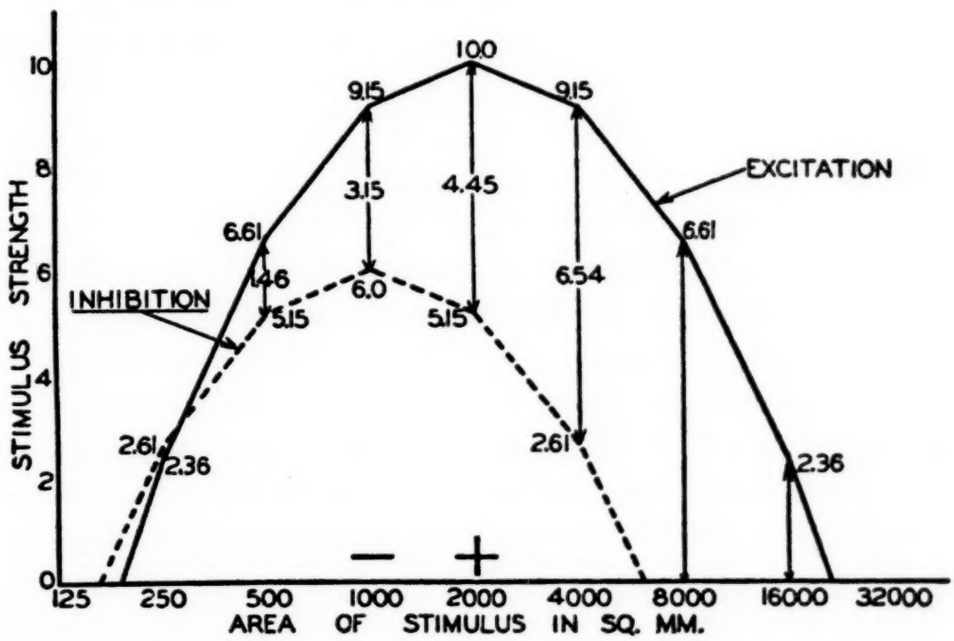


FIG. 4. Diagrammatic representation of the relationship between the hypothetical generalization curves, positive and negative, after 100 post-fixation trials of practice on the stimulus combination 1000 (−) and 2000(+).

one can it be now?" The 'behavior' following transposition is not equivalent to what goes on before. The temporal difference to which we point is a superficial one. The more significant difference in behavior is probably in terms of perceptual, emotional, and symbolical activity. If there is not equivalence of objective behavior and if the stimuli are not subjectively equivalent to the child, then can we speak of equivalent stimuli? Klüver's contribution in this connection is an invaluable one; we wish merely to point out that the term equivalence must be used only in a restricted sense,

only in so far as one is interested in describing the environment and not the organism or its behavior.

In closing this discussion, it may be pointed out that the general thesis mentioned earlier in this report, that differential transfer effects as produced through extended practice are to be understood in terms of generalization, stands as definitely supported by this study. When results from experiments of such a wide range, as conditioning of the salivary response in dogs, conditioning the galvanic skin reactions in man, transfer of maze learning in rats, transposition of discrimination learning in chimpanzees and human children seem to be subsumed under the same principle, there is hope of some unity in psychology.

IX

Summary and Conclusion

Our aim was to determine the effect of different amounts of practice of a discrimination habit upon transfer in the transposition experiment. A series of four experiments was carried through. In them, forty-five children varying in age from two and a half to six years were used as subjects. In each experiment equated groups were given different amounts of practice in a discrimination habit, then tested on transposed stimuli.

The results seem to warrant the following tentative conclusions:

1. Following a small amount of post-fixation practice of a discrimination habit, choice between transposed stimuli will be confused and unstable with frequent reverting to position habits. Following a greater amount of practice, choice upon transposition is consistently made on a relational basis. After a still greater amount of practice some subjects, at least, will make choice on an absolute basis.

2. The transition from unstable choice to a stable relative choice does not occur at the same point of practice for different stimuli. The transition occurs before thirty post-fixation trials of practice in the case of easily discriminated stimuli (visual size relata, ratio $\frac{1}{4}$). It takes place after thirty trials,

but less than one hundred in case of less easily discriminated stimuli (visual size relata, area ratio $\frac{1}{2}$).

3. The transition from a stable relative choice to an absolute choice occurs between thirty and one hundred post-fixation trials with easily discriminated stimuli (ratio $\frac{1}{4}$). It takes place after one hundred post-fixation trials, if at all, for the less easily discriminated stimuli (ratio $\frac{1}{2}$).

4. The second transition point (mentioned in conclusion number three) appears to occur somewhat earlier in case transposition is downward rather than upward.

5. The general conclusion (number one) appears to hold for weight relata as well as for visual size relata.

6. Reaction time for the first trial with transposed stimuli was markedly greater than the reaction time for last ten training trials. The increase was present in 90 percent of the cases. This increment disappeared in the course of three or four critical trials. There was some evidence to indicate that the increase was greatest at or near 'transition points.'

(Manuscript received July 7, 1938)

References

1. Bingham, H. C., Visual perception of the chick, *Behav. Mono.*, 1922, 4, pp. 104.
2. Breed, F., The development of certain instincts and habits in chicks, *Behav. Mono.*, 1911, 1, pp. 78.
3. Casteel, D., The discriminative ability of the painted turtle, *J. Anim. Behav.*, 1911, 1, 1–28.
4. Coburn, C., The behavior of the crow, Corvus Americannus Aud., *J. Anim. Behav.*, 1914, 4, 185–201.
5. Gulliksen, H., Studies of transfer of response. I. Relative versus absolute factors in the discrimination of size by the white rat, *J. Genet. Psychol.*, 1932, 40, 37–51.
6. Gundlach, R. H., and Herrington, G. B., The problem of relative and absolute transfer of discrimination, *J. Comp. Psychol.*, 1933, 16, 199–206.
7. Hovland, C. I., The generalization of conditioned responses. II. The sensory generalization of conditioned responses with varying intensities of tone.
8. Jackson, T. A., General factors in transfer of training in the white rat, *Genet. Psychol. Mono.*, 1932, 11, pp. 59.
9. Johnson, H. M., Visual pattern discrimination in the vertebrates, *J. Anim. Behav.*, 1914, 4, 319–340.
10. Jones, H. E., and Dunn, D., The configural factor in children's learning, *J. Genet. Psychol.*, 1932, 41, 3–15.
11. Kinnaman, A. J., Mental life of two macacus rhesus monkeys in captivity, *Amer. J. Psychol.*, 1901, 13, 98–148.

12. KÖHLER, W., Opische Untersuchungen am Schimpansen und am Haushuhn, *Abhandl. d. K. Preuss. Akad. d. Wissenschaften*, 1915, Phys.-math. Kl., Nr. 3.
13. KLÜVER, H., *Behavior mechanisms in monkeys*, Univ. Chicago Press, Chicago, 1933, pp. 387.
14. PAVLOV, I. P., *Conditioned reflexes*, Oxford Univ. Press, London, 1927, pp. 430.
15. RAZRAN, G. H. S., Theory of conditioning and related phenomena, *Psychol. Rev.*, 1930, **27**, 25–43.
16. SPENCE, K. W., The differential response in animals to stimuli varying within a single dimension, *Psychol. Rev.*, 1937, **44**, 430–444.
17. TAYLOR, H., A study of configurational learning, *J. Comp. Psychol.*, 1932, **13**, 19–26.
18. TAKEI, K., On visual discrimination and its learning with chickens, *Jap. J. Psychol.*, 1927, **2**, 32–87.
19. WARDEN, C. J., AND WINSLOW, C. N., The discrimination of absolute versus relative size in the ring dove, *J. Genet. Psychol.*, 1931, **39**, 328–341.
20. WARDEN, C. J., AND ROWLEY, J., The discrimination of absolute versus relative brightness in the ring dove, *J. Comp. Psychol.*, 1929, **9**, 317–337.
21. WASHBURN, M. F., AND ABBOTT, E., Experiments on the brightness value of red for the light-adapted eye of the rabbit, *J. Anim. Behav.*, 1912, **2**, 145–180.
22. WILTBANK, R. T., Transfer of training in the white rat upon various series of mazes, *Behav. Mono.*, 1919, **4**, pp. 65.

A PSYCHOLOGICAL INTERPRETATION OF THE RESULTS OF THE ZENITH RADIO EXPERIMENTS IN TELEPATHY

BY LOUIS D. GOODFELLOW

Northwestern University

Section I—Preview

Neither coincidence nor telepathy, but the natural response of an audience to secondary cues caused the 'highly successful' results of the Zenith radio experiments in telepathy. Approximately three-fourths of the audience's seventy-six attempts to receive an impression telepathically yielded results significantly different from chance expectation. Apparently, the audience was responding to definite factors—not chance. The most significant result of this study is the discovery of these factors.

This paper may be considered a study of the subtleties that influence an audience rather than a study of telepathy. It should be borne in mind *that this interpretation is based on the data available* and not on a systematic study planned by psychologists in the light of previous results.[1] Nevertheless, the high reliability and the definiteness of the tendencies uncovered due to the large number of cases make the Zenith data a valuable contribution to psychology.

Section II—The Experiment

Beginning September 26, 1937, and continuing until January 2, 1938, the Zenith Foundation (sponsored by the Zenith Radio Corporation) broadcast over a nation-wide network, every Sunday evening, a simple telepathic experiment. The program, consisting of dramatized examples of the retardation of progress through intolerance and the presentation of personal telepathic experiences, gained a huge radio audience. A few minutes of each program was devoted to the following simple telepathic test.

[1] In return for assistance which the Zenith Radio Corporation rendered the American Institute for the Deaf-blind in its research program at Northwestern University, the author interpreted the results of this experiment for the Zenith Foundation. The views here expressed do not represent the interpretations placed upon the data by the Zenith Foundation.

At a given signal, the audience was asked to determine upon which of two symbols a group of telepathic senders in the studio was concentrating. This test was repeated a number of times on each program. The radio audience, having been requested to mail their results to the radio station, responded generously. The following script illustrates the type of instructions given to the audience.

Narrator.—Tonight the experiment will make use of two of the characters on the Duke University ESP cards. A square and a circle! I'll repeat the characters to be used tonight. A circle—and a square.

Announcer.—Behind a locked door in a room near the studio are ten telepathic senders. Five are men—five are women. By means of a specially-prepared selecting machine one of the two ESP characters will be chosen at random. The ten senders will concentrate upon the character selected—they will try to send the impression of either circle or square to you. During the intervals denoted by the taps of a bell—see if the single thought in the minds of these ten senders comes through space to you. The machine will operate five times.

Narrator.—It is best to write down your impression as soon as you receive it. Do not think about it or try to reason it out. Write down your impressions in consecutive order—as rapidly as you get them. The machine is now ready to select number one.

SPIN . . . STOP . . . BELL . . . INTERVAL . . . BELL

Narrator.—That was number one. The machine will now select number two.

(ETC.)

The senders were ten students from Northwestern University. We attempted to secure people who were sympathetic toward the hypothesis of telepathy and who possessed vivid imagery. The senders worked in a small room adjacent to the studio where they were entirely unobserved. The choice of items to be sent was made by a roulette wheel and the actual selections were unknown to anyone except the senders until after the closing of the polls. The tabulation of results was made by R. L. Rolk and Company. The tabulation and checking methods used guaranteed a high degree of accuracy in the count of returns. (A sample of 9000 of the original reports submitted by radio listeners was recounted in our laboratory to secure additional data, and showed perfect accuracy in the original count.)

SECTION III—SUMMARY OF DATA

We present, in Table I, a summary of the results of fifteen tests containing in the aggregate over one million responses. Because of the great quantity of data, a table of critical ratios showing the significance of these data is more comprehensive than one showing the original data and, therefore, is given here. The original data have already been distributed to many psychology laboratories by the Zenith Foundation so that they will be available to anyone wishing to make further study.

In Table I, under the date on which the test was made, we indicate the items used for each trial on that particular

test and the significance of the audience response in terms of critical ratios. For example, looking at the September 26th test, we find that, on the first trial, 'white' was the symbol

TABLE I

SUMMARY OF CORRECT SYMBOLS AND CRITICAL RATIOS OF THE RESULTS

(A critical ratio of 3 or more is considered statistically significant. A minus before the critical ratio indicates that the audience response was below chance.)

Sept. 26 Black-White	Oct. 3 Vegetables	Oct. 10 Black-White	Oct. 17 Black-White	Oct. 24 Circle-Cross	Oct. 31 Wave-Star	Nov. 7 Square-Circle	Nov. 14 Star-Cross
1. W 11.2	Crts. −4.6	B 6.4	W 13.3	O 12.7	W 13.2	S 5.7	C −15.8
2. W 13.1	Crts. −8.4	W 7.3	B −6.0	X .40	S 5.3	S 2.4	C − 5.6
3. (B 7.8)	Corn 4.6	W 5.6	B 4.8	X 8.7	S 6.8	C .98	S − 3.0
4. W 11.8	Beans 2.7	B −1.0	W 5.5	O 2.4	W −6.2	S 11.2	C − 4.2
5. W 5.9	Beans −2.1	W 5.1	B −6.2	X 7.6	S −1.6	S −.25	S 2.5
6. B −8.9			W .14		W −5.2		
7. (B 2.2)			W 4.3		S 8.1		
Total correct responses 14.8	−3.5	10.5	6.0	14.2	7.7	9.0	−11.7
Perfect scores 10.5	—	17.9	17.7	37.3	21.8	0	13.9
Zero scores .35	—	11.8	14.5	21.0	12.5	−9.6	34.75
Number of cases 46,433	15,360	12,388	17,239	15,447	14,663	13,886	14,697

Nov. 21 Circle-Cross	Nov. 28 Cross-Star	Dec. 5 Dark-Light	Dec. 12 Heads-Tails	Dec. 19 Hot-Cold	Dec. 26 Ship-Sounds	Jan. 2 Light-Dark
1. O −17.6	S 18.0	D −37.0	T −47.6	H 16.3	H 8.5	L 24.7
2. O − 3.4	S 6.5	L 1.7	T −10.7	H −8.8	S −1.8	D −2.5
3. O 4.2	S 1.4	D 15.4	H − 9.5	H −5.7	B − .4	D 5.7
4. O − 6.2	C −7.2	L 2.8	H 4.5			D −1.7
5. O − 3.8	C −8.20	D − 7.9	H 1.10			L 4.5
6. X 4.1						
Total correct responses −9.3	4.7	−11.2	−13.7	1.0	3.6	13.7
Perfect scores −16.2	−3.6	−13.3	− 9.8	−17.8	1.7	3.2
Zero scores −15.2	−5.08	− 3.61	7.26	−16.1	3.49	−.1
Number of cases 19,178	20,367	17,548	6,644	4,499	1,605	4,521

sent, and that an above-chance number of people responded correctly, the significance of this response being expressed by the critical ratio 11.2. A minus sign before the critical ratio indicates that the audience response was below chance expectation. For example, on the sixth trial in the September

26th test, 'black' was the symbol sent, but the majority of the audience responded 'white.' Since the audience was below what should have been expected by chance, the significance of the response is expressed by a critical ratio of − 8.9.

Trials number three and seven of this first test were used as controls. Without giving the audience any indication of employing a control test, the senders on trial numbers three and seven counted rapidly to themselves instead of concentrating on 'white' or 'black.' On both these occasions, the predominant response of the audience was 'black.' This response we consider to be an attempted compensation on the part of the audience for the preceding 'white' responses, rather than an indication of personal preference. The evidence for this interpretation will be presented later.

In the row 'Total correct responses,' we have indicated the significance of the total number of correct responses in each test in terms of their deviations from expectation. Thus, in the first test of September 26th, the total number of correct responses was considerably above chance expectation, the significance being expressed by a critical ratio of 14.8.

The rows 'Perfect scores' and 'Zero scores' indicate in terms of critical ratio the significance of the number of perfect and zero scores. It will be observed throughout many of the tests that the number of perfect scores, as well as the number of zero scores, was far above chance expectation.

The last row entitled 'Number of cases' indicates the number of people sending in their guesses.

A study of these data shows some very interesting tendencies. In the first place there were very few trials on which the audience responses were what should have been expected by chance. They were generally either far above or far below chance expectation, showing that these responses were being definitely determined by certain factors other than chance. The purpose of this paper is to discover these factors.

Section IV—Response Pattern or Sequence

The key to the solution of the preponderance of extra-chance results will be found in an analysis of the pattern or

sequence which the receiver used in recording his responses. Although the importance of pattern was suspected early in the study, it was not until December 12th that we were able to secure the data required to make pattern study possible. The data from this test looked so promising that we secured samples of the original data from a number of preceding tests and recounted them in our own laboratory in order to make further pattern analysis.

The data in Table II, based on samples from six tests aggregating approximately 20,000 cases, shows the percentage of people employing each of the thirty-two possible patterns. (The star indicates the pattern actually selected in the studio in each test.) Since there are thirty-two possible patterns or permutations, we should expect in a chance situation a trifle more than three percent of the audience to use each. A glance at the column headed 'Average' shows that pattern number six was used almost thirty times as often as pattern number seventeen.

Let us look at pattern number twenty-seven which happened to be the corrected pattern for the 'light-dark' test of December 5th (DLDLD). It will be observed that, regardless of the type of test, a very small percentage of people ever recorded this sequence. The first 'light-dark' test was a failure, presumably because the senders used a non-preferred pattern. In contrast to this, let us look at pattern number fourteen which, regardless of the type of test, is highly preferred. In the 'circle-cross' test of October 24th, number fourteen was the sending pattern. This test was highly successful. The pattern sent happened to coincide with one of the audience's preferred patterns.

It should be observed that this pattern study was based on both successful and unsuccessful tests. Had the analysis been based on successful tests only, it might have been argued that certain patterns were frequently selected by the audience because of some telepathic influence. Since the audience's response patterns were the same throughout, regardless of the different patterns used by the senders, we must look to explanations other than telepathy for this phenomenon.

Before looking for an explanation of the preponderance of certain patterns, we had to determine whether the particular preferences of the radio audience were artifacts of the test

TABLE II

TABLE SHOWING PATTERNS AND THEIR FREQUENCY IN PERCENTAGE

(1 represents the first and 2 the second symbol used in each test. Thus 12121 means heads, tails, heads, tails, heads, or light, dark, light, dark, light.)

Patterns	12-5-37 Light-Dark	1-2-38 Light-Dark	12-12-37 Heads-Tails	11-7-37 Square-Circle	10-24-37 Circle-Cross	10-10-37 White-Black	Average
1. 11111	.7	.52	1.23	.07	.37	.38	.54
2. 11112	1.2	1.23	1.95	.22	.56	.76	.99
3. 11121	2.2	3.11	4.83	1.87	1.61	2.24	2.64
4. 11122	2.7	3.00	5.17	1.57	1.83	2.58	2.81
5. 11211	3.3	4.54	5.41	2.17*	2.09	2.47	3.33
6. 11212	7.6	8.61	11.06	9.04	8.82	7.86	8.83
7. 11221	8.8	8.88	10.06	6.35	6.54	5.20	7.64
8. 11222	4.4	3.84	3.92	1.35	2.98	3.64	3.36
9. 12111	2.3	2.68	2.60	1.87	1.47	1.21	2.02
10. 12112	4.6	5.79	7.63	6.20	5.01	3.11	5.39
11. 12121	3.2	3.54	3.93	2.62	2.68	1.29	2.88
12. 12122	4.2	3.70	3.65	2.77	2.84	1.75	3.15
13. 12211	7.7	7.76	7.28	6.13	6.08	4.74	6.62
14. 12212	6.5	5.97	6.84	8.30	7.51*	4.59	6.62
15. 12221	4.7	3.99*	3.51	1.87	3.54	3.19	3.47
16. 12222	1.8	1.23	.97	.37	1.13	1.14	1.11
17. 22222	.3	.59	.17	.22	.16	.34	.30
18. 22221	.4	.41	.33	.30	.38	.57	.40
19. 22212	.7	.59	.68	1.72	1.11	1.71	1.09
20. 22211	2.3	1.52	.98	1.42	1.05	2.77	1.67
21. 22122	1.7	1.09	.80	1.27	2.48	2.62	1.66
22. 22121	4.0	3.90	1.97	6.80	7.43	8.27	5.40
23. 22112	3.8	2.99	1.99	6.05	4.25	6.00	4.18
24. 22111	3.4	2.29	1.16*	1.94	2.44	2.62	2.31
25. 21222	1.1	.75	.56	1.12	1.71	1.97	1.20
26. 21221	3.1	2.56	1.79	4.33	4.47	3.49	3.29
27. 21212	1.0*	1.18	1.19	2.84	1.87	.68	1.46
28. 21211	1.6	2.02	1.67	3.06	2.96	3.95	2.54
29. 21122	3.7	3.47	2.18	4.71	5.96	5.69	4.29
30. 21121	4.2	4.77	2.72	6.88	5.05	6.64*	5.04
31. 21112	2.2	2.73	1.21	3.96	2.68	5.31	3.02
32. 21111	.6	.75	.47	.82	.93	1.56	.86
Standard error	.55	.26	.23	.48	.25	.34	.12
Size of universe	17,548	4,403	5,687	13,886	15,447	12,388	
Size of sample	1,000	4,403	5,687	1,341	5,033	2,635	

* Pattern chosen for sending.

situation, or whether they represented a genuine tendency on the part of the audience to follow certain particular patterns.

In exploratory tests conducted at Northwestern University among classes and fraternity groups, we found that the same patterns predominated regardless of the type of test. The test forms were distributed to groups with instructions that they were to be filled out at a later time and in private so as to avoid being influenced by other persons either through marginal sensory cues or telepathy. The data collected on each of the four following forms showed no significant variation from the preferred patterns found among the radio audience.

1. Please write in each space below one of the words, 'Black' or 'White.'

1._____ 2._____ 3._____ 4._____ 5._____

2. Please fill each space below with either the plus (+) sign or the minus (−) sign.

1._____ 2._____ 3._____ 4._____ 5._____

3. MENTAL COIN-TOSSING TEST.

Imagine that you are tossing a coin (Don't actually toss one), and indicate in the spaces below which side came up in each mental flip of the coin.

1._____ 2._____ 3._____ 4._____ 5._____

4. PREDICTING ZENITH PATTERN.

The Zenith test for next Sunday will consist of a five trial test using the symbols 'star' and 'cross.' Guess which symbol will be chosen for each trial, and record your prediction below.

1._____ 2._____ 3._____ 4._____ 5._____

An interesting verification of this result comes from Stanford University in which Dr. John Kenedy in a study of the patterns used in longer series of 'calls,' found that the same patterns predominate. We shall assume then that the cause of the predominance of certain patterns lies in the individual rather than in the objective environment. (One exception to this finding is discussed below in connection with the effects of pattern preference on psycho-physical research.)

In order to make a more careful study of the effect of pattern per se, we have combined each pattern with its reci-

procal. (By reciprocal, we mean a pattern in which each symbol has been replaced by its alternative.) For example, DLDLD has been combined with LDLDL since the two patterns are identical. Whether an individual uses a pattern or its reciprocal depends entirely upon his choice of the first symbol. A great preference or predisposition on the part of

TABLE III

PATTERNS AND THEIR RECIPROCALS COMBINED TO INDICATE TRUE EFFECT OF PATTERN PREFERENCE

(11111 represents the frequency of 11111 plus 22222, etc.)

Patterns	12-5-37 Dark-Light	1-2-38 Dark-Light	12-12-37 Heads-Tails	11-7-37 Square-Circle	10-24-37 Circle-Cross	10-10-37 Black-White	Average	
1. 11111	1.0	1.11	1.40	.29	.53	.72	.84	
2. 11112	1.6	1.64	2.28	.52	.94	1.33	1.39	
3. 11121	2.9	3.70	5.51	3.59	2.72	3.95	3.73	
4. 11122	5.0	4.52	6.15	2.99	2.88	5.35	4.48	
								10.44
5. 11211	5.0	5.63	6.21	3.44*	4.57	5.09	4.99	
6. 11212	11.6	12.51	13.03	15.84	16.25	16.13	14.23	
7. 11221	12.6	11.87	12.05	12.40	10.79	11.20	11.82	
8. 11222	7.8	6.13	5.08*	3.29	5.42	6.26	5.66	
								36.70
9. 12111	3.4	3.43	3.16	2.99	3.18	3.18	3.22	
10. 12112	7.7	8.35	9.42	10.53	9.48	6.60	8.68	
11. 12121	4.2*	4.72	5.12	5.46	4.55	1.97	4.34	
12. 12122	5.8	5.72	5.32	5.83	5.80	5.70	5.70	
								21.94
13. 12211	11.4	11.23	9.46	10.84	12.04	10.43	10.90	
14. 12212	10.7	10.74	9.56	15.16	12.56*	11.23*	11.66	
15. 12221	6.9	6.72*	4.72	5.83	6.22	8.50	6.48	
16. 12222	2.4	1.98	1.44	1.19	2.06	2.70	1.95	
								30.99
Size of universe	17,548	4,403	5,687	13,886	15,447	12,388		
Size of sample	1,000	4,403	5,687	1,341	5,033	2,635		

* Pattern chosen for sending.

an audience to use one symbol rather than another will cause an unusually large response on the one pattern and an unusually low response on its reciprocal, or vice versa. Combining the patterns and their reciprocals eliminates this phenomenon, and thus gives us a picture of the true effect of pattern, per se. Table III shows the data of Table II grouped in this manner.

In Table IV, we have arranged the patterns and their reciprocals in the order of the frequency used by the audience.

Our hypothesis explaining this great audience preference for some patterns is based on a study of the symmetry occurring in the pattern. Our assumption is that the audience, probably unconsciously, avoided a highly symmetrical or orderly pattern. This may seem strange at first, since one would expect a person to prefer a symmetrical pattern on æsthetic grounds. The situation here was quite different. The receiver knew that the correct pattern was a result of chance

TABLE IV

TABLE SHOWING PATTERNS, THEIR SYMMETRIES, FREQUENCY USED BY AUDIENCE, AND TIMES CHOSEN AS SENDING PATTERN

Patterns	Symmetries of Groups						Average Frequency	Times Used
	2	3	4	5	1–3	Total		
11212.....	1	2	0	0	0	3	14.23%	1
11221.....	2	0	1	0	0	3	11.82%	0
12212.....	1	1	1	0	0	3	11.66%	4
12211.....	2	0	1	0	0	3	10.90%	0
							48.61	
12112.....	1	1	1	0	1	4	8.68%	0
12221.....	2	1	0	1	0	4	6.48%	1
12122.....	1	2	0	0	1	4	5.70%	0
11222.....	3	1	0	0	0	4	5.66%	1
							26.52	
11211.....	2	1	0	1	0	4	4.99%	1
11122.....	3	1	0	0	1	5	4.48%	1
12121.....	0	3	0	1	1	5	4.34%	1
11121.....	2	2	0	0	1	5	3.73%	0
							17.54	
12111.....	2	2	0	0	1	5	3.22%	0
12222.....	3	2	1	0	0	6	1.95%	0
11112.....	3	2	1	0	1	7	1.39%	0
11111.....	4	3	2	1	1	11	.84%	1
							7.40	

selection. Since symmetry suggested an orderly arrangement, the receiver avoided a symmetrical or orderly pattern.

In Table IV, we have listed the number of symmetries occurring in the different patterns. Under '2' in the column headed 'Symmetry of Groups,' we have listed the number of symmetrical pairs occurring in each pattern. For example, the second pattern listed in Table IV contains two symmetrical pairs, items one and two, and items three and four. Under '3' in the column headed 'Symmetry of Groups,' we have listed the number of symmetrical triads. Thus, the first

pattern contains two symmetrical triads: items two, three, and four; and items three, four, and five. Each forms a symmetrical triad. The column headed '1–3' has been added to give a little extra weight to the first three items since we have found them particularly important in psycho-physics. In both laboratory experiments with psychophysical judgments and in the Zenith data, we have observed a very marked disposition on the part of our observers to make their third response different from the first, thus avoiding a symmetrical triad. The explanation of this observed fact is not clear, but possible reasons for it are discussed below. Furthermore, the extra weighting of the first triad helps to differentiate between patterns which would otherwise have the same weight. Under 'Total' the number of symmetries occurring in each pattern as defined above is given. It will be observed that the least symmetrical patterns were the ones most frequently used by the audience in recording responses, and the most symmetrical patterns were the ones least used.

Possibly symmetry is a poor term. We might consider this avoidance of symmetry as a tendency to change one's response due to one's previous responses. In recording his second trial, the listener had only one previous response to consider. The question was, should it be the same or different, and the data show that for approximately fifty-three percent of people it was different. There is a widespread misconception among laymen that the occurrence of a chance event, such as the tossing of 'heads,' reduces the probability of the same event occurring the second time.

When the receiver came to the recording of his third response, his problem was more involved because his tendency to make trials two and three different was complicated by his effort to make the first group of three appear disorderly. On the fourth and fifth trials, the tendencies operating became more and more complicated.

We do not mean to imply that a person actually assumed this rational approach. We merely suggest that these attitudes and habits of thinking formed a background which, in the long run, produced these trends. Regardless of how we

describe this phenomenon, or what its true cause be, the fact that observers tended to follow certain sequences and avoid others must be considered in evaluating this type of data.

This tendency of an observer to follow a definite pattern is not limited to experiments in telepathy. A study[2] of patterns used by observers in psycho-physical measurements shows that there are definite preferences in the various sequences used by our observers, but the particular patterns used by the observers can be altered by the instructions, and experimental set-up. Other investigators[3,4] have been interested in the same phenomenon although they considered it from the standpoint of the influence of previous responses on one's judgments.

Table V shows the dates on which the various patterns were chosen by the selecting machine. It will be observed that patterns preferred by the audience happened to be chosen by the studio for sending more often than the less preferred ones. This does not mean that the chances of getting a high score were increased. It means, merely, that the chances of getting the correct pattern were increased, and, therefore, one's chance of getting either a very high score or a very low score (depending on one's choice of the first symbol) were increased. A person choosing the incorrect symbol on the first trial, but following the correct sequence would be wrong on each succeeding trial. Because preferred sequences happened to be chosen frequently in the studio for the sending pattern, there were a great many more perfect scores and a great many more zero scores than should have been expected (see Table I). A study of Table IV shows that 45 percent of the time, the sending pattern happened to be one of the four most preferred by the audience, whereas only 9 percent of the time the sending pattern was one of the four consistently avoided by the audience.

[2] Goodfellow, L. D., Ranseen, Emil, and Heine, R., The significance for psychometry of tendencies to follow certain patterns in recording responses (awaiting publication).

[3] Preston, M. G., Contrast effects and psychophysical judgments. *Amer. J. Psychol.*, 1936, **48**, 389–402.

[4] Arons, L., and Irwin, F., Equal weights and psychophysical judgment. *J. Exper. Psychol.*, 1932, **15**, 733–51.

However, the excess of low scores did not cancel the effect of the high scores, for as was shown above, when all the data were averaged the results were still significantly above chance. This was due to the fact that the audience started with the correct symbol more often than with the incorrect one. Table I shows that in ten tests the audience started with the correct symbol, and in five tests with the incorrect one. Since the first response determined the succeeding ones,

TABLE V

SHOWING PATTERNS, THEIR FREQUENCY, AND THE DATES OF TESTS IN WHICH THEY WERE USED AS SENDING PATTERNS

Patterns	Average Frequency	Quartile Totals	Dates
11212.....	14.23%		Nov. 14
11221.....	11.82%		
12212.....	11.66%		Oct. 10, 17, 24, 31
12211.....	10.90%	48.61%	
12112.....	8.68%		
12221.....	6.48%		Jan. 2
12122.....	5.70%		
11222.....	5.66%	26.52%	Dec. 12
11211.....	4.99%		Nov. 7
11122.....	4.48%		Nov. 28
12121.....	4.34%		Dec. 5
11121.....	3.73%	17.54%	
12111.....	3.22%		
12222.....	1.95%		
11112.....	1.39%		
11111.....	.84%	7.40%	Nov. 21

the results of each test hinged on the response to the first trial. Furthermore, since the audience response on trial number one was very decided—due to the large extra-chance factors—a study of the determiners of the first response became very important.

SECTION V—HYPOTHESES CONCERNING THE DETERMINERS OF THE FIRST RESPONSE

Before considering the possibility that telepathy was the explanation of the overwhelming number of correct beginnings made by the audience on ten of the fifteen tests, let us look at two other possibilities. First, we might assume that an

individual's first choice would be determined by his 'personal preference' for the symbols being used. Secondly, there might have been objective factors in the experimental set-up which caused the results obtained.

Let us now look at the personal preference hypothesis. This preference might have been due (1) to a greater familiarity with one of the symbols, (2) to a particular meaning or significance attached to one of the symbols, or (3) to mere ease in recording.

TABLE VI

SHOWING PERSONAL PREFERENCE FOR TEST ITEMS BASED ON EXPLORATORY EXPERIMENTS INVOLVING NO INTENTIONAL TELEPATHY

Black-White	Vegetables	Coin Tossing	Light-Dark	ESP Characters	
				Using Symbols	Using Words
W 50.13%	Carrots 18.04	*H* 58.5	L 54.1	Circle 24.5	22.7
	Beans 16.31			Cross 20.1	19.6
B 49.87%	Peas 26.04	*T* 41.5	D 45.9	Square 20.8	19.5
	Corn 20.65			Star 18.7	19.2
	Beets 18.96			Wavy-lines 16.8	19.0
Standard error .51%	.4%	1.5%	1.3%	.57%	.52%
No. of cases 9810	9810	1105	1460	4850	5975

By means of exploratory tests we have attempted to establish the personal preference for the various symbols. These data were collected from four principal sources: (1) university classes, (2) a civic organization of a very small middle western town, (3) customers at the Atlantic and Pacific, and National Tea Stores, and (4) groups of clerks in Chicago business houses. The data were collected in the most part by students from the university. Instructions were presented to the observers in writing to secure uniformity. We have no assurance that our instructions were free from leading statements, so that part of the preferences indicated in the table might conceivably be an artifact of the instructions. The possibility that some telepathic influences from the experimenter or dominant person in the group might have affected the results was discounted, since the results

of a group-test did not differ significantly from the results of individual test conducted by an experimenter ignorant of the test he was administering.

In Table VI, we show the percentage of responses for each item and the total number of responses made. In the case of the test for preferences for the ESP characters, the data were collected in two different forms. One group of people was asked to record their response by drawing the actual symbol, while another group was asked to write the word, as was done in some of the Zenith tests. A test of this type is open to

TABLE VII

SHOWING PERSONAL PREFERENCE AS INDICATED BY FREQUENCY OF SYMBOL ON FIRST TRIAL OF EACH EXPLORATORY TEST

Black-White	Coin Tossing	Dark-Light	ESP Characters	
			Using Symbols	Using Words
W 52.1	*H* 78.2	L 66.4	Circle 34.6	32.4
			Cross 21.3	20.8
B 47.9	*T* 21.8	D 33.6	Square 20.6	19.6
			Star 12.7	13.9
			Wavy-lines 11.8	13.3
Standard error 1.6%	3.3%	2.9%	2.9%	2.6%
Number of cases 981	221	292	194	239

serious criticism. We suspect that many individuals attempted to compensate for their first choice by naming a different one on the following responses. In fact, many persons named three of one symbol and two of the alternative in the five-trial two-alternative tests. This would indicate the direction of their preference, but the actual percentages become rather meaningless. In some respects it seemed that the response on the first trial would be a more reliable index of personal preference. Those data are given in Table VII.

The data from Tables VI and VII indicate a decided preference for 'light' over 'dark,' 'heads' over 'tails,' 'circle' over the other ESP characters, and 'cross' and 'square' over 'star' and 'wavy-line.' The figures on the other items (*e.g.*, black-white) lack statistical significance: the deviation from

the percentage expected by chance being less than three times the standard error.

Since our concept 'personal preference' seems inadequate to explain the predominant response by the audience on the first trial of many tests (cp. Tables I and VII), we turn to a second hypothesis, namely, that subtle suggestions, or cues, accidentally occurring in the script might predispose the audience to begin the test with one of the symbols rather than the other.

As experimentists, we are always concerned over the suggestive nature of our instructions. Most persons recognize the great influence of subtle suggestions on human behavior, but many of us do not appear to recognize fully the effects on psychological experiments. In our laboratory, we have demonstrated [5] that a threshold may be altered by a suggestion in the instructions. Since the radio participants had no obvious criterion upon which to base their response, it does not seem unreasonable to suspect them of using any slight cue or suggestion as a basis for their response.

A study of phonographic recordings of the broadcast by a group of fifteen to twenty professional persons known as our critics, to determine whether there might have been any objective factors in the preceding script suggesting one of the symbols, reveals some interesting possibilities. To illustrate these possible leading suggestions, let us look at the 'star-cross' tests of November 14th and 28th. The script contained many subtle suggestions of 'star,' our critics believed. The predominant response of the audience on trial number one of both tests was 'star' notwithstanding the fact that 'cross' was more often chosen on our exploratory test. The instructions for the tests are given below and the alleged suggestions of 'star' italicized.

NARRATOR.—Two of the characters on the Duke University ESP cards will be used in this experiment. So jot down on your paper. Cross and *Star*. (Emphasized by announcer.) I'll repeat. A cross and a *star*.

ANNOUNCER.—*Alone* in a room *high above the streets of Chicago* are ten *especially selected* ('S' sound) telepathic *senders—five* men and *five* women (Five—characteristic

[5] Goodfellow, L. D., Factors affecting threshold measurements. (To be published soon.)

points of star not cross). The *selecting* machine to be used has been prepared to *select* at random one of those two ESP *characters*—either a *star* or a cross. You will hear the machine operate—a bell will ring—and those ten *senders* will concentrate upon the *character selected* each time. The machine will operate *five* times. Each *selection* must be and will be either a cross or a *star*. There will be no blanks. During the *five second* intervals *see* if the *single* thought in the minds of those *senders* comes to you.

NARRATOR.—Write down your impressions in consecutive order. Do not hesitate or try to reason these out. Write them down as *soon* as you get them. The machine will *select* number one.

It must be borne in mind that these alleged suggestions italicized in the instructions are merely the opinions of our group of critics. It might be difficult to explain for example, why the phrase, 'high above the streets of Chicago,' is suggestive of star rather than cross since the cross (on a church) is also frequently seen high above the streets. However, fourteen out of eighteen persons, when asked which symbol the phrase suggested to them, replied 'star.' Regardless of the identity of the individual items suggesting 'star,' the total effect of the instructions seemed to favor 'star' rather than 'cross' as is shown in Table VIII. The plausibility of this hypothesis is increased when we consider that, to account for the results actually obtained, only thirteen percent of the audience need have responded to these cues. In some of the other tests, a response from an even smaller percentage of the audience was sufficient to produce the observed results.

Our critics suggested at least two simple plans for objectively testing our hypothesis, but they were never used. For example, the 'cross-star' test could have been repeated using instructions from which the alleged suggestions of 'star' had been deleted, and possibly subtle suggestions of 'cross' added. Another plan offered was to conduct a test on two different programs using 'star' and 'circle.' By using the October 24th instructions which presumably favored 'circle' for one test, and the November 14th instructions which were allegedly weighted for 'star' for the second test, we would have had a nice objective test of the hypothesis of the predisposing effect of the instructions.

After studying the instructions given the audience, our critics pointed out certain cues which they unanimously agreed might have influenced the audience. We were interested in

finding out the suggestive material which could be cited by persons who did not know the actual results of the tests, so requested twenty-eight persons to read the scripts carefully and indicate which, if either, of the symbols was suggested by the preceding instruction. Although there is no unanimity of opinion, this analysis supports the previous study made by our critics. Since our critics (in most cases) had the advantage of hearing a recording of the actual program, and therefore had more information upon which to base their judgments, we shall use their evaluation in our analysis.

TABLE VIII

SHOWING JUDGMENTS ON THE SUGGESTIVE VALUE OF THE DIFFERENT SCRIPTS

The first column gives the number of persons considering the script suggestive of the first symbol, the second column, the number considering it suggestive of the last symbol, and the third column, those considering the scripts of neutral value. The last column gives the judgment of our critics after hearing phonographic recordings of the broadcasts.

Test	Suggestive of First Symbol	Suggestive of Second Symbol	Neutral	Critic Conferences
Sept. 26 (Black-White).....	4	11	13	White
Oct. 10 (Black-White).....	4	15	9	White
Oct. 17 (Black-White).....	5	17	6	White
Oct. 31 (Wave-Star).....	14	6	8	Wave
Nov. 7 (Square-Circle).....	10	6	12	Square
Nov. 14–28 (Star-Cross).....	22	2	4	Star

For the reader's interest, we show Table VIII, in which the suggestive value of each test as judged by independent raters is given.

For some time we have been interested in the suggestive value of instructions in psychometric studies. It has been demonstrated a number of times in various psychological laboratories, that the instructions may change the results of an experiment. As a result of our studies in sensitivity, we have long suspected that when a judgment becomes difficult, an observer will grasp upon the slightest cue available.

Recent experimental[6] evidence tends to support this hypothesis. In an experiment in which observers were asked

[6] Goodfellow, L. D., and Heine, Ralph, The effect of incidental factors on threshold measurements. *Proc. Ill. Acad. Science*, 1938.

at a given signal to place their fingers upon a vibrator, and determine whether or not it was vibrating, we introduced a slight change in the ready signal, and found that an observer, if properly motivated, would use this slight difference in the ready signal as the basis of his judgment.

Another study [7] comes from Dr. Worchester of the University of Nebraska. Students asked to write down a number between one and twenty-five recorded a predominance of odd

TABLE IX

TABLE SHOWING THE PERSONAL PREFERENCE AND THE SUGGESTIONS ALLEGEDLY OPERATING IN EACH TEST, AND THE SYMBOL CHOSEN BY SENDERS AND RESPONSE MADE BY THE AUDIENCE ON THE FIRST TRIAL FOR EACH TEST

(A question mark (?) means that the preference as measured by our exploratory tests lacks statistical significance.)

Date	Test	Personal Preference	Suggested by Instructions	Selection Made on First Trial	
				By Studio Machine	By Audience
Sept. 26.....	Black-White	White?	White	White	White
Oct. 10.....	Black-White	White?	White	Black	Black
Oct. 17.....	Black-White	White?	White	White	White
Oct. 24.....	Circle-Cross	Circle	Circle	Circle	Circle
Oct. 31.....	Wave-Star	Star?	Wave	Wave	Wave
Nov. 7.....	Square-Circle	Circle	Square	Square	Square
Nov. 14.....	Star-Cross	Cross	Star	Cross	Star
Nov. 21.....	Circle-Cross	Circle	Neutral	Circle	Cross
Nov. 28.....	Star-Cross	Cross	Star	Star	Star
Dec. 5.....	Dark-Light	Light	No data	Dark	Light
Dec. 12.....	Coin Tossing	Heads	No data	Tails	Heads
Jan. 2.....	Light-Dark	Light	No data	Light	Light

numbers. When the instructions were changed, asking for a number between two and twenty-six, the even numbers were more frequently used.

Regardless of the lack of objective data concerning the influence of subtle suggestion on results, there appear to be adequate data from similar situations to warrant a suspicion that cues in the instructional script are significant in these experiments.

Table IX summarizes the data on personal preference and the alleged suggestive cues.

[7] Worchester, D. A., The effect of instructions on guessing, reported before the Mid-Western Psychological Association at Madison, Wisconsin, in May, 1938.

The last column gives the predominating response made by the audience. Let us assume that this was determined by one or more of the items indicated by the first three columns: namely, preference, suggestion, or the studio choice (telepathy). Which of these three factors was most highly correlated with the actual result?

The first test does not help us to answer this question since all three factors (preference, suggestion, and telepathy) have the same effect, that is, to produce the response 'white.' However, the second test presents a different situation. In spite of the suggestions of 'white' in the instructions, the audience chose 'black'—the correct symbol. This might have made a good case for telepathy if it were not for subsequent findings. In this program and also in the 'circle-cross' test of November 21st, certain suggestive material (discussed below) was inserted just before the test instructions. In both cases, the audience response correlated with these suggestions. On the November 21st test, this suggestive material pointed to the incorrect response, and the audience responded accordingly. Apparently, this suggestive material was stronger than telepathy. Of course, there is no positive proof that these suggestions were the determining factors, but in both cases the audience response correlated with them, notwithstanding that on the November 21st test, the audience responded oppositely to what would have been expected, had telepathy been operating.

The leading suggestions referred to are these. On October 10th, it happened that, immediately preceding the test, Mr. Gutzom Borglum reported his experience conducting Helen Keller through European art galleries. It may have been possible that this emotional suggestion of blindness, darkness, and black carried over to predispose part of the audience to start off with the symbol 'black.' (A response to the suggestion of 'black' was required from less than six percent of the radio audience to produce the results actually obtained.)

One factor which we have observed in psychological experiments, and which appears to function here, is that suggestions or cues, to be effective, must be subtle. When

they become obvious enough for the subject to recognize them as cues, they will be resisted. Compare this, for example, with the old detective story technique of making the correct cue obvious, and thus misleading the reader. This process seemed to happen on the 'star-cross' tests of November 14th and 28th among a group of students from Northwestern University taking the test. Having previously discussed with us the possible existence of cues, they were aware of this psychological principle, and, therefore, were alert to the suggestions of 'star.' They resisted the suggestions, starting off both tests with the symbol 'cross.' The rest of the radio audience started with 'star' on each occasion. On the first 'star-cross' test (November 14th) 'cross' was the chosen symbol for sending, and on November 28th, 'star' was chosen.

The 'circle-cross' test of November 21st was particularly interesting in this respect. On this one occasion, the script was sent to us ahead of time for the purpose of being freed of cues. Most of the persons in our special group helped rewrite the instructions and were satisfied that they were as free from suggestion as possible. On this test 'circle' was chosen by the studio, and the results for both the country as a whole and our special University group were below chance expectation, notwithstanding the fact that their personal preference for 'circle' should have caused a correct response on the first trial. In other words, both the University group and the public resisted their personal preference for 'circle' and the telepathic attempt of our senders, starting the test with the incorrect symbol 'cross.' Since we had made a special effort to delete from the instructions any suggestive lines, the reason for this peculiar phenomenon, we believe, lies in the immediately preceding announcement (inserted at the last moment) of the correct answers for the 'star-cross' test of the previous week. The symbol 'cross' was given a prominent place in the stimulus background of the audience because it had been the correct symbol on three of the five trials. (This was the only test in the series in which this type of suggestive script was possible.) It is interesting that our special University group, having satisfied themselves that

there were no cues in the script, responded to this suggestion of 'cross.' Being taken off guard, they, as well as the general public, followed this lead and were, therefore, wrong.

Thus, it would seem that the response of the audience on the first trial in each test was not a matter of chance, but was determined by subtle uncontrollable factors. In the long run one would expect these factors to balance out, that is, to predispose the audience toward the incorrect response just as often as toward the correct one. From this approach, each test became a single case in which these variable factors either oppose or facilitate the correct response. Fifiteen tests are far too few cases to permit chance to operate sufficiently to remove the effect of the determiners of the first response. As a matter of fact, the first response of the audience, although always definite, was correct on ten tests and incorrect on five—a proportion well within the realms of chance. Since the first response determined the success or failure of the whole test, and it in turn appears to have depended upon a number of varying factors, any crucial evidence for telepathy must show a significant number of tests in which the first response is correct. The Zenith series of tests contained only fifteen—a number far too low for statistical treatment. Any implication substantiating telepathy is an obvious artifact of the experimental set-up.

Section VI—Evidence against Telepathy

The following facts to be discussed in this section clearly indicate that telepathy was not a factor in these particular tests. Negative results of this sort carry no implication as to the validity of the extra-sensory hypothesis in general, except that they suggest a possible source of error which should be carefully checked.

1. The audience was consistent in following certain patterns or sequences, regardless of whether the pattern was right or wrong.

2. On three tests which were repeated, the audience response did not change in spite of the fact that a different pattern was chosen for sending. On the fourth repeated test,

a change in audience orientation changed the audience response on each of the first four trials. Furthermore, this change made the audience wrong.

3. Persons selected because of high scores on the first ten programs did not differ significantly from the rest of the radio audience on succeeding tests.

4. The statistical significance of the whole series of tests is due to the large number of cases in the first test. When the first test is discounted, the results of the whole series are reduced to chance. The above-chance average for the whole series rests on this one heavily-weighted successful test.

5. Twenty-five special groups in various laboratories throughout the country made a chance score when they tried to receive impressions from the radio audience concentrating simultaneously.

6. After determining the factors operating, scientists successfully predicted the outcome of six tests before any returns were received.

1. *Study of Audience Response Patterns*

As was pointed out above in the discussion of patterns, the audience was highly consistent in the pattern or sequence of symbols it followed regardless of whether the pattern was right or wrong. The fact that the audience did not change its pattern of response when a different pattern was sent shows that the factors causing the radio audience to use certain patterns or sequences was at least stronger than telepathy.

Had the slightest amount of telepathy been operating in these tests, the audience should have used the correct pattern a little more frequently than they used this pattern in their responses to other tests. In no case did this happen. The fact that the senders were concentrating upon a particular pattern did not cause any increase in the frequency with which that particular pattern was used by the audience (see Table III).

2. *Comparison for Repeated Tests*

In Table X we compare the responses on four tests which were repeated at a later date, using a different sequence of

symbols. It is interesting that there were very few significant changes in audience response when the same kind of test was repeated, notwithstanding the fact that a different order of symbols was sent. Of the eight changes which did take place two lead to correct responses and four lead to incorrect responses. The other two changes resulted in a chance response.

TABLE X

COMPARISON OF REPEATED TESTS, SHOWING THE SYMBOL CHOSEN EACH TIME BY THE AUDIENCE

The question mark (?) indicates that the audience response was merely what could be expected by chance. A minus sign (−) means the audience's choice was incorrect.

	White-Black		Circle-Cross		Star-Cross		Dark-Light	
	9/26	10/17	10/24	11/21	11/14	11/28	12/5	1/2
1	W	W	O	−X	−S	S	−L	L
2	W	−W	?	−X	−S	S	?	?
3	B	B	X	O	−C	?	D	D
4	W	W	?	−X	−S	−S	?	?
5	W	−W	X	−X	?	−S	−L	L
6	−W	?						
7	?	W						

The comparison of the two 'circle-cross' tests is interesting because the audience response changed on each of the first four trials. This, we believe, was due to a complete reversal of audience orientation due to the subtle suggestions in the preceding script discussed above. When the test was repeated, a suggestion of the incorrect symbol was unwittingly introduced. This caused the audience to reverse its first response. The different orientation resulted in a change in the succeeding responses.

The comparison of the 'light-dark' test is also interesting. When this test was repeated on January 2nd, we were fortunate in having one of Dr. Rhine's experienced senders do the concentrating instead of the ten Northwestern University students. A glance at Table X shows that the audience response on the two tests was identical on each trial, without exception. The radio audience started with the response 'light,' followed by 'chance,' 'dark,' 'chance,' and 'light' re-

sponses in each test. Regardless of the sending, the audience responded in exactly the same manner. The second time, however, the studio pattern coincided with the audience's pattern, and this test was highly successful. Why? If telepathy caused the above-chance results on this occasion, what caused the above-chance response on this particular pattern in the other tests when the studio was attempting to send different patterns?

This interesting set of data (Table X) indicates two things: first, that the audience responded to definite factors, not chance, and secondly, that telepathy is not one of these factors.

3. *Special Receivers*

One very curious fact seemed to be that there was an unduly large group of people who were making high scores.

TABLE XI

REPORT ON SPECIAL RECEIVERS

	Perfect Score			Percent Correct		
	Entire Audience	Special Receivers	Critical Ratio	Entire Audience	Special Receivers	Critical Ratio
Dec. 19	3.8%	2.3%	−.7	50.0%	51.0%	.33
Dec. 26	18.2%	22.2%	1.3	35.8%	38.7%	1.4
Jan. 2	3.9%	3.4%	−.5	54.6%	54.5%	−.09

These persons lived in all sections of the country, and represented both men and women, both young and old. On December 15th, letters were mailed to 375 persons who had scored high in the ten previous tests, requesting them to mail their results directly to the secretary of the Zenith Foundation for separate tabulation on the three following tests. The scores from these persons were tabulated separately. Table XI shows that these special receivers did not differ from the ordinary receiver either in the number of correct responses or the number of perfect scores. The reason for the unusually high number of perfect scores, we attribute to the fact that, as was shown above, a favorite sequence was accidentally used by the senders a large number of times. This caused the

excess of perfect scores. It also caused an excess of zero scores.

The Zenith Foundation was very much interested in having us make some individual telepathic tests on some of these persons who were highly successful in the radio experiments. The individual, out of thousands participating in the radio experiment, who had the highest score in the radio tests was unable to come to Chicago because of his health, so the second best person was secured. As we did not feel that we had either the technique or the proper attitude to conduct a fair telepathic test, we secured one of Dr. Rhine's former assistants to conduct it for us. Out of 1100 calls using the ESP cards, the average number of hits per run by this prize receiver and experienced sender was 4.9 (expected by chance — 5.0).

4. *Significance of Whole Series*

The great statistical significance of the whole series of tests which the Zenith Foundation announced [8] as having odds of one to 10,000,000,000,000,000,000 against chance was no exaggeration. However, this was due to the great number of responses on the first test. When this one test was discounted, the statistical significance of the entire series was reduced to chance. Table XII shows the results on each test. The results for the whole series, with the first test eliminated, are indicated. The statistical significance of the whole series becomes less impressive when one realizes that a single heavily-weighted 'hit' is responsible for it.

Furthermore, five of the fifteen tests average significantly below chance, indicating definitely that extra-chance factors were producing the results. In at least these five tests, it must be admitted that these factors outweigh the effects of telepathy. In order to allow these factors to cancel each other, we must consider each test as a single case. This makes the number of cases for consideration in this respect not over a million, but merely fifteen (the number of tests).

[8] What Well-known Scientists Say about Telepathy (distributed by the Zenith Foundation), page 17.

TABLE XII

SUMMARY SHOWING THE NUMBER OF CORRECT RESPONSES AND ITS SIGNIFICANCE FOR EACH TEST

Date	Correct Responses	Expected by Chance	Difference	Critical Ratio
Sept. 26 (White-Black).....	118,246	114,703	3,543	14.8
Oct. 3 (Vegetables) *.....	14,172	14,560	−388	−3.5
Oct. 10 (White-Black).....	31,969	30,672	1,297	10.5
Oct. 17 (White-Black).....	61,204	60,162	1,042	6.0
Oct. 24 (Circle-Cross).....	40,499	38,521	1,978	14.2
Oct. 31 (Line-Star).....	52,389	51,156	1,233	7.7
Nov. 7 (Square-Circle)....	35,836	34,652	1,184	9.0
Nov. 14 (Star-Cross).....	35,089	36,670	−1,581	−11.7
Nov. 21 (Circle-Cross).....	55,700	57,272	−1,572	−9.3
Nov. 28 (Star-Cross).....	51,245	50,503	742	4.7
Dec. 5 (Dark-Light).....	42,002	43,663	−1,661	−11.2
Dec. 12 (Heads-Tails).....	13,680	16,144	−2,464	−13.7
Dec. 19 (Hot-Cold).....	6,751	6,692	59	1.0
Dec. 26 (Ship Sounds) *....	1,722	1,605	117	3.6
Jan. 2 (Light-Dark).....	12,269	11,240	1,029	13.7
Total.....	556,879	552,050	4,829	9.2
Total excluding first test....	438,633	437,347	1,286	2.7

* Excluded from totals because these data are not comparable.

Ten 'hits' and five 'misses' can hardly be offered as evidence of the operation of telepathy. Considering these facts, we should place the odds at one to two, not one to 10,000,000,000,000,000,000.

5. *Special Audience-Sending Test*[9]

On February 13th arrangements were made with twenty-five groups of people throughout the country to act as receivers while the whole radio audience did the sending. These groups were locked in rooms where they could not hear the radio, but could receive a signal from the person conducting the test at the time to concentrate. The symbol 'black' or 'white' was selected in the studio, the announcement made to the audience, and the audience asked to concentrate while these special groups tried to receive. The results show that the number of perfect scores was far below chance-expectation. The number of correct responses on each individual trial, and on the test as a whole, was chance, as is shown in Table XIII.

[9] For further discussion see Fernberger, S. W., Extra-sensory Perception or Instructions?, *Jour. Exper. Psychol.* 1938, **22**, 602–7.

TABLE XIII

SHOWING THE RESULTS OF THE SPECIAL AUDIENCE-SENDING TEST OF FEBRUARY 13TH

	Perfect Scores	Correct Responses
Actual results	.58%	50.30%
Expected by chance	3.12%	50.00%
Number of cases	339	1695

Further analysis is difficult since we do not know the exact instructions given the various groups by the persons conducting the experiment. However, the resultant effect was zero, since there was no evidence of extra-chance factors operating in this test. The data showing the patterns used by each participant are not available.

6. *The Prediction of Results*

After the first six tests, an analysis of the data seemed to point to several definite factors producing the results submitted by the radio audience. On the assumption that if we knew the causal factors operating, we should be able to predict the result, we proceeded to test our assumptions by predicting the outcome of the tests.

At the time, we did not have the pattern data available, so that our predictive formula was derived from data collected from our University campus groups. It was evident from these data that the outcome of the whole test was largely determined by the response on trial number one. The first step in predicting the outcome of a test was to determine experimentally the strength of the factors determining the first trial. This we found by reading the instructions to small groups of people and asking them what their first response would be. Knowing this we expected the second response to be the same, but with a smaller deviation from chance. The third response, we discovered, would be the opposite and with about the same deviation from chance as the second. The last two responses were insignificant compared with numbers one and three, but were generally the same as number one. In short, the audience aggregate pattern was 11211, 1 representing the favorite or suggested symbol and 2 representing the alternative one.

The proportion of the audience using the symbol called for by this formula depended on the strength of the personal preference or leading instructions involved. For example, on the November 7th test, there was not a very decided preference on trial one, and therefore, only a slight reversal on number three. On December 5th and 12th, a very decided response on trial number one was followed by a marked reversal on trial number three. On a basis of these principles, we predicted the outcome of six tests before any of the results were received from the radio audience. This, to us, was evidence that we had found the most important factors operating in the experiment, and that telepathy was not one of them. This was evident from the fact that the tests of November 14th and December 5th and 12th, all of which gave results far below chance, were predicted in advance.

The accuracy of our predictions depended upon our success in evaluating the determiners of the first response. For this reason we probably would have failed on the 'circle-cross' test of November 21st, had we tried to predict its outcome because we did not recognize the suggestive material which was inserted just before the test. (It will be recalled that this is the test from which we attempted to delete all suggestive material.) Assuming that we had non-leading instructions, we would naturally have expected the audience to start with the preferred symbol 'circle,' as they did on the October 24th test. 'Circle' was the correct symbol on this occasion (November 21st), but the audience started with the symbol 'cross,' due, we believe, to the suggestion of 'cross' made immediately preceding the test.

Section VII—Premature Investigations

The data to be reported here will appear to be premature since these particular analyses were made before an adequate technique was developed. However, they are reported in hope that someone may make something of them, and to avoid being guilty of selecting the data for this report.

1. *Analysis by Geographical Location*

Table XIV shows the results of four geographical groups, New York City, Chicago, California, and rural sections of the Midwest, analyzed both for perfect scores and total correct responses. In all cases the differences were not greater than those to be expected by chance. The announcement made over the air to the effect that the data showed that time and space had no effect on telepathy was unfortunate, since there

TABLE XIV

ANALYSIS BY GEOGRAPHICAL LOCATION

(*In terms of percentage of correct responses*)

	Nov. 14	Nov. 21	Nov. 28	Dec. 5	Dec. 12
New York	46.77	48.94	49.75	48.34	44.04
Chicago	46.28	48.76	51.39	48.16	42.37
California		48.45	50.68	47.71	39.07
Rural sections of Midwest	47.46	48.83	50.72	48.49	41.63
Entire country	47.84	48.63	50.73	49.10	42.37

(*In terms of percentage of perfect scores*)

	Nov. 14	Nov. 21	Nov. 28	Dec. 5	Dec. 12
New York	4.28	.20	2.72	1.62	.94
Chicago	5.09	0.00	3.32	1.75	1.66
California		.20	2.25	1.30	.42
Rural sections of Midwest	4.19	.06	2.93	1.52	.44
Entire country	5.12	.10	2.68	1.38	1.04

is no evidence that these tests were measuring telepathy. The announcement that scientists had authorized this statement was in error.

2. *Analysis by Age*

Table XV shows the averages for three age groups, under 20, 20–50, and over 50, in terms both of percentage of perfect scores and percentage of correct responses. In no case were the differences greater than what could have been expected by chance. Here again the announcement made over the air to the effect that the group under 20 had a higher percentage of correct responses was unfortunate, although literally accurate.

TABLE XV

COMPARISON OF THREE AGE GROUPS

	Percentage of Perfect Scores			Percentage of Correct Responses		
	Under 20	20–50	Over 50	Under 20	20–50	Over 50
Dec. 5	1.06	1.42	1.70	48.57	47.75	47.84
Dec. 19	3.40	3.90	3.60	52.30	49.21	50.20
Dec. 26	16.70	17.80	19.50	33.81	35.54	37.11
Jan. 2	5.00	3.90	3.90	54.24	53.87	55.29

3. *Response from Persons Living in Steel Buildings*

On December 12th the sending was done from a frame building, and the audience was asked to indicate whether they lived in a frame or steel building. If telepathic communication depends on electromagnetic waves, a steel building would act as a screen, thus preventing communication. The results in Table XVI compare the two groups as to the number of perfect scores and number of correct responses. Neither is significant.

TABLE XVI

TABLE COMPARING RESPONSES FROM PERSONS LIVING IN STEEL AND FRAME BUILDINGS

	Frame	No. of Cases	Steel	No. of Cases	Difference	Sigma of Difference	Critical Ratio
Perfect scores	1.09%	5149	.82%	607	.27%	.74	.36
Correct responses	41.59%	5149	40.53%	607	1.06%	.96	1.10

SECTION VIII—SEX DIFFERENCES

Table XVII shows the results of men and women analyzed separately. The comparison of correct responses shows no significant results. However, in comparing the number of perfect scores, we find that on three occasions the women have significantly surpassed the men in the number of perfect scores. Women also had a larger proportion of zero scores. This difference is readily accounted for on a basis of the following pattern analysis. Table XVIII shows pattern preferences for men and women separately. The most striking fact here is that there are fewer women than men using the highly symmetrical patterns. In other words, the female is more

ardent in her avoidance of symmetry. We might look at it another way, namely, a tendency on the part of men to take a little more of a chance on a symmetrical pattern. Whatever may be the cause, this tendency of the female groups to avoid the symmetrical patterns will account for the observed difference in men and women on the percentage of perfect and zero scores.

TABLE XVII

COMPARISON OF MEN AND WOMEN

Date	Correct Responses			Perfect Scores		
	Men	Women	Critical Ratio of Difference	Men	Women	Critical Ratio of Difference
Oct. 10	51.075	51.492	−.44	5.517	6.605	−3.2
Oct. 17	50.473	50.615	−.18	1.807	2.160	−2.5
Oct. 24	51.516	53.280	−2.13	7.82	9.16	−4.5
Oct. 31	51.150	51.094	.07	2.22	2.68	−3.1
Nov. 7	52.056	51.237	.95	3.418	2.795	2.0
Nov. 14	48.162	47.509	.74	4.94	5.48	−1.7
Nov. 28	50.4026	50.2453	.22	2.77	2.59	−.69

The minus sign before a critical ratio in above tables indicates that women were superior.

TABLE XVIII

TABLE COMPARING PATTERN PREFERENCE OF MEN AND WOMEN

(Based on samples of two tests. N = 5500)

Patterns	Men	Women
11212	16.08	17.07
12212	11.62	13.33
11221	10.15	13.04
12211	11.04	11.97
12112	8.08	8.43
12221	7.58	6.83
12122	4.46	6.81
11222	6.50	4.06
11211	5.35	3.92
11122	4.11	3.93
12121	3.28	2.68
11121	3.57	2.51
12111	4.05	2.20
12222	2.42	2.01
11112	1.16	.73
11111	.56	.51

Conclusions

An analysis of over a million responses from the radio audience, reveals the operation of marked extra-chance factors. The two most important of these factors are (1) the pattern or sequence used by individuals in recording their guesses, and (2) the set or predisposing influence of subtle suggestions found in the test instructions. When these factors are discounted, it becomes unnecessary to postulate telepathy to explain the results. The Zenith data resembles many of the data reported by telepathy exponents in that the results can not be explained on a basis of chance. However, as has been demonstrated, telepathy is not the only alternative even when such factors as sensory cues, errors due to r‹ ›rding methods, and the selection of favorable data are eliminated. It would be interesting to know what an analysis of the type made in this paper would reveal in such data.

(Manuscript received July 10, 1938)

THE INFLUENCE OF DELAYED INSTRUCTIONS TO LEARN UPON HUMAN PERFORMANCE [1]

BY E. H. PORTER, JR.

University of Oregon

Blodgett (1) and Tolman and Honzik (3) have shown that the delayed introduction of reward greatly influences the maze performance of rats during the non-rewarded trials, but leaves the final amount of learning relatively uninfluenced. The general procedure was to run a group of rats through a maze for several days with no food at the end box. A control group always ran to food. Before the introduction of reward, the non-rewarded group gave no evidence that they had learned the maze. However, after the reward was introduced, the performance of the experimental group soon equalled that of the control group.

These experiments are interpreted by Tolman (2) as disproving the Law of Effect. According to the Law of Effect good effects resulting from a response cause the connection between that situation and that response to be stamped in, and bad effects resulting from a response cause the connection to be stamped out. For the non-rewarded groups in the experiments referred to above, there existed no such differential effects during the initial trials. Or, at least it may be said that the differential effects were much greater for the control groups than for the experimental groups. Nevertheless, the experimental groups evidently learned during the initial trials.

This paper presents an attempt to determine whether comparable conditions would lead to comparable results with human subjects. A previous study with somewhat comparable conditions is that made by Twitmyer (4) who had two

[1] I wish to express my appreciation to Dr. C. S. Hall and to Dr. C. L. Huffaker for many critical comments during the course of the investigation and to Dr. R. W. Leeper and Dr. L. F. Beck for their aid in the preparation of this manuscript.

groups of laboratory students administer a maze to each other in turn. The group that learned the maze first required much more to learn than did the subjects who first served as experimenters, although these latter subjects were unaware that they subsequently would be expected to learn the maze themselves. While Twitmyer's study offers evidence that human subjects will learn more quickly after an initial exposure to the material without instructions to learn, the study lacks suitable control of time factors. One group was exposed to the maze for a greater period of time than the other group.

Experimental Procedure[2]

Experiment I.—This experiment involved the learning of a set of twenty nonsense syllables exposed for two seconds each. To avoid leading the subjects in the experimental group to suspect that they would eventually be asked to reproduce the nonsense syllables they were first given a short free-association test which served to have paper at their chairs. They were then told that they would be shown a group of nonsense syllables which were to be used later in an experiment with a different group. They were asked to pay close attention to the syllables as they were being exposed and to raise their hands if any syllable reminded them of a meaningful word. The syllables were then exposed five consecutive times, after which the subjects were instructed to record on the paper available all of the syllables they could recall. After giving instructions to learn as many of the syllables as possible five subsequent exposures were made and a second recall period given. A control group was given two periods of five exposures each followed by recall periods. This group, however, was given instructions to learn the syllables before each period of exposure. The experimental group was composed of sixty-one subjects, the control group of forty-three. All were elementary psychology students.

Experiment II.—The material to be learned consisted of a series of fifteen technical psychological terms, unfamiliar to the subjects. The experimental group was told that a list of psychological terms would be read and that they were to record on slips of paper provided for the purpose a classification for each word as to whether they thought it referred to a behavior characteristic, a physical characteristic or to a piece of apparatus. The list was read three times, after which the subjects were instructed to record all of the terms they could recall. Two subsequent sets of three presentations were given followed by recall periods and preceded by instructions to learn. The words were read to the control group nine times in all, a recall period being held at the end of every third reading. This group was given instructions to learn from the outset. The experimental group contained sixty-one subjects, the control group twenty-nine.

Experiment III.—In this experiment two problem-solving situations were used as the learning material. Each problem consisted in a simple riddle presented in masked

[2] For a more detailed description and the actual materials used the reader is referred to the writer's thesis, "The Influence of a Period of 'Incidental' Learning upon Subsequent Learning with Intention," deposited at the University of Oregon Library at Eugene.

form by being incorporated into the body of a short narrative passage mimeographed on an 11″ × 8½″ sheet of paper. The experimental groups were instructed to outline the passage as quickly and as well as possible. At the end of ten minutes they were stopped and instructed to seek out and solve the riddle presented in the passage. The control groups were instructed at the outset to seek the riddle and try to solve it.

The subjects were sophomores and juniors from two different high schools. At one school the juniors were given problem A, and the sophomores problem B. At the other school the juniors received problem B, and the sophomores problem A.

RESULTS [3]

Experiment I.—Figure 1 shows the positions of the experimental and control groups in respect to the number of syllables recalled at each testing.

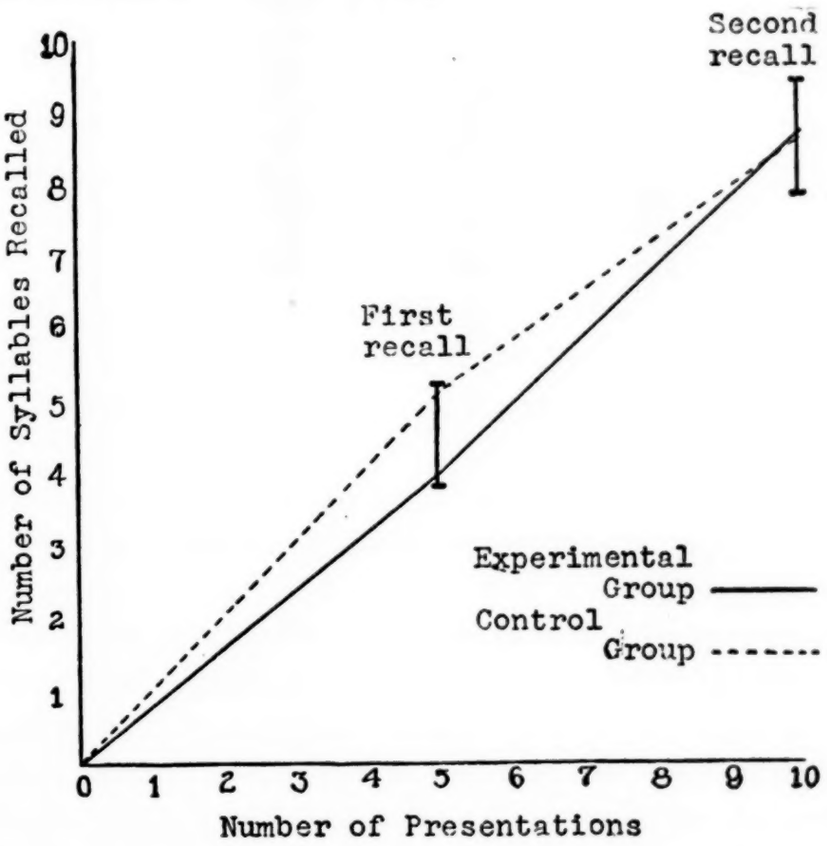


FIG. 1. Performance curves for experimental and control groups in experiment I. Vertical bars indicate 3 $\sigma_{diff.}$.

A comparison of the two performance curves shows the control group leading at the end of the initial period by a mean difference of 1.1, the significance of which is evidenced by a critical ratio of 2.5. At the second testing, however, the

[3] Since the material presented in this paper is largely corroborative of results of animal studies, only the main results are given.

experimental group leads by a mean difference of 0.9, a difference of no significance. The point to be emphasized is the ultimate equality of the performance of the two groups.

Experiment II.—In Fig. 2 the relative performance of the two groups at each recall is presented. The achievement of the control group is taken as 100 percent at each point. The relative amount achieved by the experimental group is in this way more clearly illustrated.

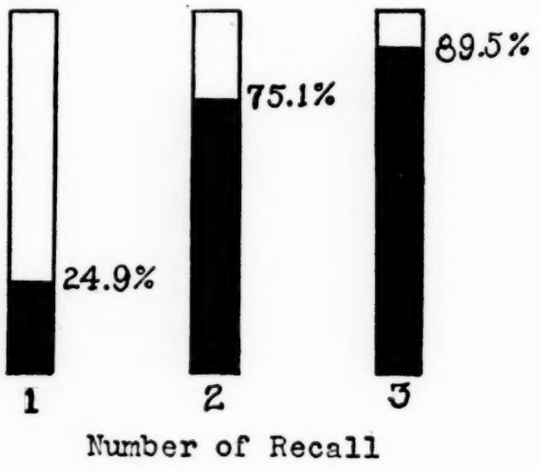


FIG. 2. Relative achievement of experimental and control groups in experiment II.

Figure 2 graphically illustrates the rapid rise of the experimental group to a level of performance almost equal to that of the control group.

Experiment III.—In Table I is presented the reliability of the mean-differences of the experimental and control groups in problem-solving time. The time values given are for the time spent in problem-solving only. The total time spent by the experimental groups is in each case 10 minutes more than the values given.

The results show the tendency of the procedure to produce positive mean-differences which indicates that an initial period of outlining activity is effective in reducing problem-solving time. The effect shows up more clearly in the more advanced groups. This point is strikingly illustrated by the greatly superior performance of the group of juniors who worked on problem A.

TABLE I

RELIABILITY OF MEAN-DIFFERENCES IN PROBLEM-SOLVING TIME

Problem	N	Group	M	σ_M	CR
A	14 (jr.)	Exp.	12.40	2.38	12.7
		Cont.	29.64	1.19	
	21 (so.)	Exp.	27.81	4.19	0.1
		Cont.	27.62	2.97	
B	22 (jr.)	Exp.	4.09	.75	4.4
		Cont.	15.27	1.98	
	18 (so.)	Exp.	10.11	1.47	2.8
		Cont.	15.33	2.32	

DISCUSSION OF RESULTS

The results from the three experiments may best be treated in terms of comparable performance periods: (1) the initial period, in which the experimental groups were given instructions other than instructions to learn and (2) the post period, the period of performance subsequent to the introduction of instructions to learn.

Initial Period.—The results in general indicate that the experimental groups learned during the initial period even though no instructions to learn were given and even though none of the subjects reported either an expectation to recall the presented material or the exercise of an intention to learn. It is assumed that the reports of the subjects are sufficient evidence that intention to learn was not present. It is further assumed that the absence of intention to learn is a condition comparable to the absence of food in the maze situation for animals, a condition which favors greater achievement by the control groups. Where measures of efficiency are directly obtainable the control groups without exception show greater efficiency accompanying instructions to learn during the initial period.

Post Period.—The performance records indicate that, after the introduction of instructions to learn, the learning rate of the experimental groups is increased to a degree greater than their learning rate during the initial period and greater

than the learning rate of the control groups for the chronologically corresponding periods of presentation. It is necessarily assumed that the later performance records are comparably adequate measures of the amounts learned during the past period.

Since the experimental groups were at lower levels of achievement at the end of the initial period, it might be argued that the more rapid rate of learning demonstrated during subsequent periods of presentation is a function primarily of greater ease of acquisition at lower levels. That this is not necessarily so is demonstrated by the fact that the control groups required more trials to achieve a comparable amount. This analysis is made by estimating on the graphs the number of trials required by the control groups to achieve an amount equal to that achieved by the experimental groups during the post period or part thereof.

A possible explanation for this increased learning rate may be that the experimental groups were more highly motivated during the post period than were the control groups for the corresponding period, although there is no definite proof that motivation has any effect other than leading the organism to react toward its environment, thus providing some of the essential conditions for the operation of the mechanisms of learning. It cannot be assumed that the presence of a motive assures the operation of the mechanisms of learning. Under these circumstances it would be more suitable to assume that the performance of the experimental groups during the post period expresses a greater degree of utilization of material learned. An increase in motivation may have resulted from the novelty of the situation, or it may be that these subjects felt challenged when they were told that they were competing with a group that, from the start, had had instructions to learn. All groups were informed that they were competing with others: the control groups were told at the outset, the experimental groups at the time of the introduction of instructions to learn. It is also possible that the contro, subjects became progressively less motivated as the experiment continued.

In the case of the problem-solving experiment the time savings may be accounted for on the basis of the experimental subjects' having been led to do things which they would have had to do anyway before solving the problem. The superior performance of the groups of juniors requires still further explanation. It is conceivable that where problem-situations have been less often met and methods of attack are less familiar that factors tending to increase the rate of solution would be of less influence than where such situations are common and methods of attack more familiar.

Interpretation of Results

In the experiments of Blodgett and Tolman and Honzik the animals of the experimental group were allowed to run in the maze for a period of time during which no food was placed in the end box. At the same time control animals ran to food. According to the Law of Effect, as it is usually stated, the situation was so ordered that the control group of animals would derive 'good' effects from taking the 'correct' path and 'bad' effects from taking the 'incorrect' path. This did not hold true for the experimental group of animals, however, as there was little or nothing in the situation to make one path more or less suitable than another. Later, when food was introduced at the end box for the experimental group of animals, the situation was changed so that it might be said that responses would produce 'good' or 'bad' effects according to whether they led to or away from food.

The significance of the experiments lies in the fact that although the 'good' and 'bad' effects were not present in the experimental situation until late in the task, the final achievement was just as great as though they had been present throughout. This result would not be expected on the basis of the Law of Effect in its customary formulation.

In the present experiments it is assumed that according to the Law of Effect the instructions to learn served to introduce 'good' and 'bad' effects into the situation, allowing the responses involved in memorizing to be followed by 'good' effects and other, non-relevant responses to be followed by

'bad' effects. If this assumption is accepted, it may then be argued that the differential effects existed throughout the entire task for the subjects of the control groups but existed for the subjects of the experimental groups only after the introduction of the instructions to learn. It is on the basis of the above stated assumption and the reasoning which follows that the experiments reported here are considered comparable with the animal experiments to which references have been made.

Summary and Conclusions

The subjects in the experimental groups were led to react to nonsense syllables, technical terms, and riddles, first, without intentions to learn and later with instructions to learn. The final achievement was equivalent to the achievement of control subjects who received the same material for the same amount of time but who from the outset had been instructed to learn.

On the basis of the assumption that the introduction of instructions to learn served to institute or establish differential effects of responses, it may be concluded that the study agrees with Tolman's statement that "The Law of Effect, if valid at all, is only slightly so, since the amount of improvement acquired without differential effect is practically as great as that acquired with differential effect" (1932, p. 344).

(Manuscript received July 8, 1938)

References

1. Blodgett, H. C., The effect of the introduction of reward upon the maze performance of rats. *Univ. Calif. Pub. Psychol.*, 1929, 2, 8, 113–134.
2. Tolman, E. C., *Purposive Behavior in Animals and Men*, New York, The Century Company, 1932.
3. Tolman, E. C., and Honzik, C. H., Introduction and removal of reward, and maze performance in rats. *Univ. Calif. Pub. Psychol.*, 1930, 4, 257–275.
4. Twitmyer, E. M., Visual guidance in motor learning. *Amer. J. Psychol.*, 1931, 43, 165–187.

CONDITIONING IN THE WHITE RAT. IV. THE CONDITIONED LID REFLEX

BY BARBARA HUGHES AND HAROLD SCHLOSBERG

Brown University

This paper describes the development of the conditioned lid reflex in the white rat and compares it with other conditioned reflexes in the rat and with the same reflex in other animals.

Earlier papers in this series (**6**, **8**, **9**) have described experiments in which shock to the tail or fore leg served as the unconditioned stimulus for conditioning in the white rat. Perhaps the most outstanding result of these experiments was the difficulty encountered in establishing a conditioned withdrawal of the to-be-shocked member. Diffuse responses, as breathing changes, squeals, and general tonic shifts, were conditioned in one or two sessions (15–30 double stimulations), but precise fore-leg flexion became stably conditioned in only one out of some 30 rats, and then it required 10 sessions.

Does this mean that no specific responses can be conditioned in the white rat while it is restrained in a holder? Or was the chief source of difficulty the unconditioned response employed, namely, withdrawal of a member from shock? To answer these questions, it seemed desirable to attempt to condition some response not dependent upon shock. The lid reflex seemed well suited to the purpose, for it is a sharp, precise response. It is easily elicited by a puff of air to the cornea, and reasonably specific, rather than general. This conditioned lid reflex offers another advantage in that Hilgard and his associates have already accumulated data on the response in man (**1**, **2**), monkey (**4**), and dog (**3**). The present study will add another species to their comparative program.

METHOD [1]

After considerable preliminary exploration, satisfactory apparatus was developed. The animal was fastened in a holder, and placed in a sound-resisting box. The holder was a modified form of one used in the earlier studies. A laced sleeve and four leg thongs held the rat astride a block of wood, and an additional thong over the occiput held the neck and lower jaw firmly pressed against the block. To hold the head reasonably immobile it was necessary to add a padded flat chain over the lower jaw, and an adjustable pressure pad against the skull just back of the eyes. Rats became adapted to the holder and ceased frequent struggling after 4–6 half-hour pre-training periods.

Stimuli were controlled by a compound pendulum outside the sound-resisting box. Depression of a key released the pendulum and simultaneously started the high-frequency buzzer which served as the conditioned stimulus. The pendulum terminated the buzz 800 ms. later, by opening a knife switch. At the same time it struck a rubber bulb, forcing a puff of air through appropriate tubing to the rat's cornea.

Preliminary experimentation made it clear that some device would be necessary to keep the eyelid open between stimulations. This suggested the use of a light mechanical recording system, so arranged that it would exert a slight tension against lid closure. Accordingly a light muscle lever was set up with an 8-inch strand of thin rubber to furnish the tension. A thread passed from the lever through glass bearings in the top of the sound-resisting box, to the eyelid of the rat. Attempts to fasten the thread to the eyelid with adhesives were not successful. Finally a loop of surgical silk was sewn through the lid (under anæsthesia) to which the recording thread could be attached. This method, although scarcely ideal, was successful if care was taken in adjusting the light cellophane writing tip of the muscle lever. The movements of this lever and of magnetic markers for buzzer and time ($\frac{1}{4}$ sec.) were recorded on an extension kymograph.

Before conditioning was attempted, each rat was placed in the holder for 4 or more pre-training sessions. After the rat had learned to remain quietly in the holder, ether was administered and a loop was sewn through the lid. After a day of rest, the rat was put through a program patterned after Hilgard's and Marquis' (3). On the first day (buzz adaptation) responses to 60 presentations of the buzz, distributed over one hour, were recorded. On day 2 (puff adaptation) 50 responses to the air puff were recorded, together with 5 responses to the buzz at the beginning and end of the hour. Conditioning sessions, starting on the third day, consisted of 5 presentations of the buzz, 50 buzz-puff, and 5 more buzz. After 4 to 7 conditioning sessions were obtained, 2 sessions were devoted entirely to extinction by unreinforced presentations of the buzz.

THE DEVELOPMENT OF THE CONDITIONED RESPONSE

There are at least three responses that may be involved in any conditioning experiment, the unconditioned and conditioned responses, and the original response to the conditioned stimulus. In the present experiment the unconditioned response was of course the lid reflex to a puff of air on the cornea.

[1] A more detailed description may be found in the original thesis, on file in the Library of Brown University. We are indebted to Mr. F. L. Reinwald for assistance with the apparatus.

This reflex, when tested on the preliminary 'puff adaptation' day, was sharp, and of short latency. The speed of recording did not permit fine latency measurements, but it was adequate to show that the median latency of the unconditioned response is around 50 ms.

The original response to the buzzer, as measured on the 'buzz adaptation' day, was extremely slight. From 3 percent to 7 percent of the records for that day showed responses, and these were so variable in latency and form that they should probably be considered as incidental winks rather than as

TABLE I

DEVELOPMENT AND EXTINCTION OF THE CONDITIONED RESPONSE IN TERMS OF FREQUENCY AND MEDIAN LATENCY OF ANTICIPATORY RESPONSES

	Rat	Conditioning Sessions							Extinction Sessions	
		1	2	3	4	5	6	7	1	2
Frequency percent	C	27	67	47	53	50	60	73	64	54
	D	45	30	32	38	63	71		57	30
	B	47	62	81	52				73	44
	M	19	70	59	90				72	53
	Mean	35	60	55	58	57	65		67	45
Latency ms.	C	775	415	255	435	320	295	160	135	205
	D	778	700	745	570	740	275		215	298
	B	560	470	400	320				400	285
	M	190	215	215	170				157	165
	Mean	575	450	404	375				227	262

reflex responses to sound. The data from the 'buzz adaptation' day show that 10 percent or more responses in a session may be taken as an indication of conditioning.

The conditioned response appeared during every one of the conditioning sessions for each rat. Its form, amplitude, and latency varied with the stage of conditioning. At first the conditioned response was slow and small. As conditioning proceeded, it closely approached the unconditioned reflex in form and amplitude, but only to a limited extent in latency. In Table I median latencies are reported for each rat for each day, together with corresponding data on the frequency with which the conditioned reflex appeared. An examination of

the table will show that there is an inverse relationship between latency and frequency. When plotted on a scatter diagram the relationship is seen to be roughly linear, with some spread, but with only two sessions badly out of place (Day 5, Rat D; Day 1, Rat M). The rank difference correlation between frequency and latency is $-.43 \pm 13$. This happens to be the exact value of the σ for a similar correlation in dogs (3) and essentially the same as that reported for monkeys (4). But the distribution of latencies in the rat covers a much wider range than the ones reported for man, dog and monkey. This is largely due to the length of the period of delay. In man, dog and monkey the onset of the puff masked any conditioned responses of latencies longer than 375–400 ms. In the rat, latencies as long as 800 ms. were possible. The exclusion of long-latency responses apparently modifies the distribution markedly, so that they cannot be compared directly. But an examination of the trend of our data suggests that a well-developed conditioned lid reflex in rats would have a latency of 150 ms.–200 ms. This is within the range of latencies reported for dog, monkey and man.

As we have seen, frequency and latency are two of the criteria one may use in indicating the development of the conditioned response We have presented both in Table I and have suggested that they show about the same picture. The frequency data probably present the simplest and most straightforward account, so our discussion will be carried on largely in these terms. Frequency data have been plotted in Fig. 1 for the reader's convenience. Referring to this figure, it may be seen that the response developed rapidly at first, appearing in response to 35 percent of the buzzes during the first day. It reached an average frequency of 60 percent during the second session, and maintained this value through the remaining 2–5 conditioning periods.[2] When these results

[2] The number of conditioning and extinction sessions was partly controlled by the condition of the loop sewn through the lid. The records of two additional animals were discarded because the loop pulled through the lid. The results from the discarded animals are presented in the original thesis. They differ in no significant way from the ones reported here.

are compared with those reported by Hilgard and Campbell for man (**1**, **2**), and by Hilgard and Marquis (**3**, **4**) for dog

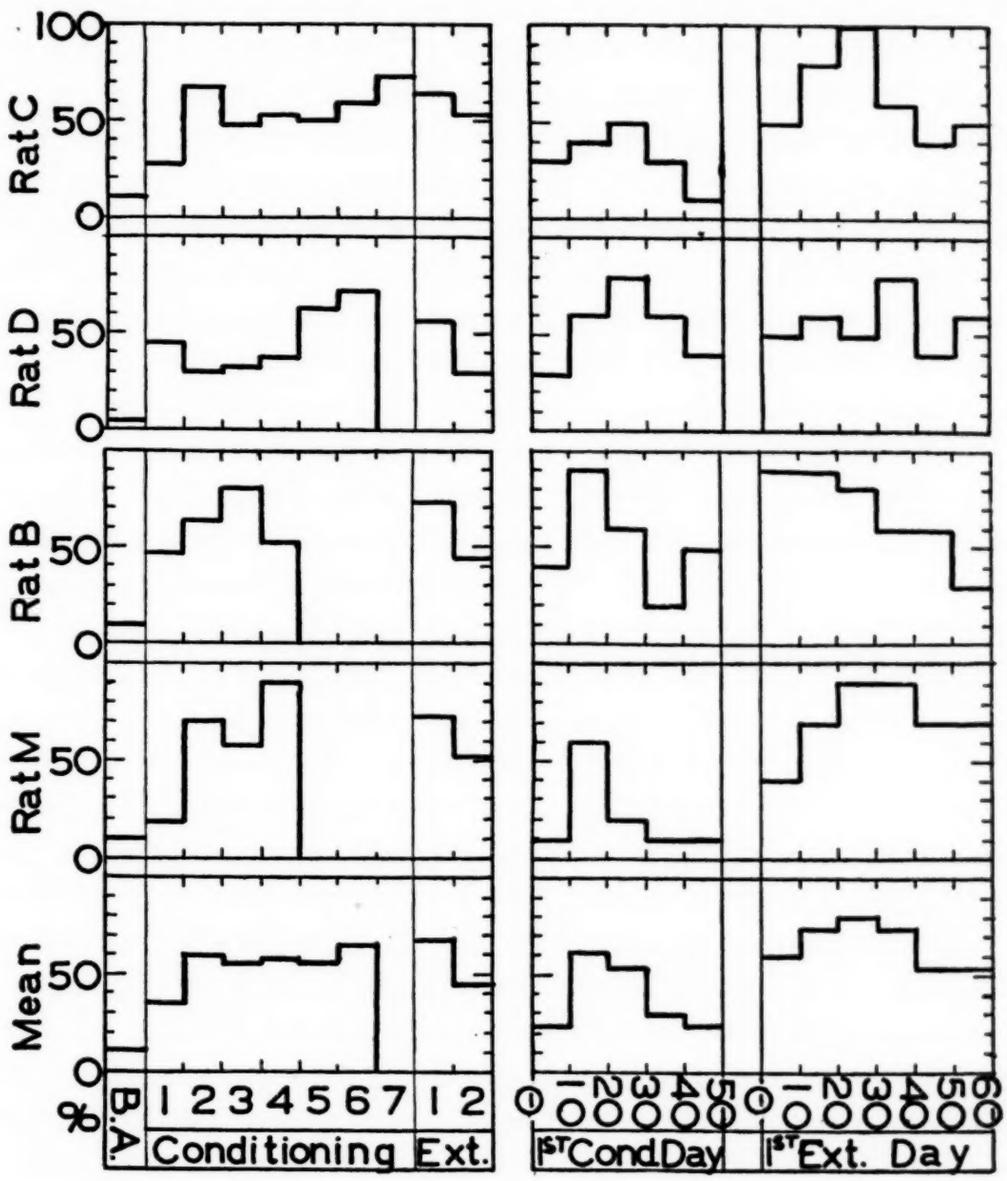


FIG. 1.

FIG. 2.

FIG. 1. Daily percentages of conditioned lid responses to buzz which anticipated the shock. The successive vertical columns are: B.A. = Buzz Adaptation; Conditioning Sessions 1–7; Extinction Sessions 1 and 2.

FIG. 2. Percentages of conditioned lid responses appearing in successive groups of 10 trials. The first conditioning session is on the left; the first extinction session is on the right.

and monkey, some interesting similarities and differences appear. The results with the rat are apparently within the wide range obtained from man. But the rats reached a higher

mean frequency on the first day than did either the dogs or monkeys. Only one out of four dogs showed any conditioned response on the first conditioning day, and on that day made but 2 responses out of 50 chances. The monkeys showed no more responses on this day than on the 'light adaptation' day. The rats, however, improved less rapidly than did dogs or monkeys. The four dogs caught up to the rats during the second session, and four out of five monkeys passed the rats on the fifth day. Thereafter, the four monkeys and the dogs reached and maintained a frequency between 80 percent and 90 percent, which is clearly above the 60 percent level attained by the rats.

It is quite probable that the superiority of rat over the dog and monkey during the early conditioning sessions is spurious. It will be remembered that the buzz preceded the puff to the rat's cornea by 800 ms. But the period between light and puff with the dog and monkey was 375 ms. and 400 ms. respectively. If in our experiments the puff had come 400 ms. after the onset of the buzz it would have masked any conditioned responses with a latency longer than 400 ms. This would have eliminated about half of the conditioned responses tabulated in this paper, especially those in the early sessions.

As we have indicated above, the rats were markedly inferior to dogs and monkeys in the final frequency level reached. This seems to be due to the failure of the rat to continue to improve after the second or third session. Some insight into this failure to improve can be gained from the daily records. In Fig. 2 (left side) the first day for each animal is broken down into successive groups of 10 stimulations, and the number of conditioned responses occurring in each group of 10 trials is plotted as a percentage. Similar curves were obtained on successive conditioning days. The highest incidence occurred during the second or third group of 10 trials, and after that there was a marked falling off. Both Liddell (**7**) and Schlosberg (**8**, **9**) have reported that there is a 'warming up' period at the beginning of each session during the establishment of a conditioned leg flexion, so it is not surprizing to find one here. Further, Schlosberg has reported

that the conditioned response to shock develops in fewer sessions when sessions are short (25 stimulations) rather than long (200 stimulations). Thus the later portion of a long session seemed to have a negative effect in the leg withdrawal experiments just as it does here. Similar daily curves have been reported by Hilgard and Marquis for the dog and monkey. They attribute the shape of the curve to 'antagonistic effects of practice (increment) and of continuous work (decrement)' (**3**, p. 56). Still more recently Hovland (**5**) has indicated that 'inhibition of reinforcement' exists, and is greatest when trials are massed. To these general negative factors, whatever they may be, there must be added another unfavorable factor specific to this experiment, namely, the effect of the loop sewed through the lid. The fundamental purpose of this loop was to exert tension against lid closure, and in so doing it probably interfered with conditioning. Of course one might expect any irritation caused by the loop to increase toward the end of the session. Apparently, then, a number of factors led to a falling off of the conditioned response during the second half of each session. This falling off was greater than it had been with the dog or monkey, and resulted in poorer conditioning in the rat. The obvious solution of the difficulty is to halve the number of stimuli per session.

Extinction

After from 4 to 7 days of conditioning sessions, each rat was subjected to two days of extinction sessions. In each of these extinction sessions the responses to 60 isolated buzzes distributed over one hour were recorded. On the right side of Fig. 2 we have plotted the frequency of responses obtained for each group of 10 trials of the first extinction session. It may be seen that 2 of the rats (C, M) showed a clear gain in frequency during the first half of the session, followed by a loss during the second half. A third rat (D) showed some suggestion of this early gain, while the remaining rat (B) fell off consistently. The rise in frequency during the early stages of extinction might be considered as further evidence

that 'inhibition of reinforcement' (Hovland, 5) had accumulated from too great massing of stimuli during the conditioning sessions. But any such inhibition should have been lost during the 23 hours that elapsed between conditioning and extinction sessions. It is more probable that the initial rise is another example of the 'warming up' also found during the training sessions. At any rate, all four rats showed a clear loss on the second of the two extinction days (Fig. 1). Extinction of this response in the rat seems comparable in speed to that found in man and dog, but much more rapid than in the monkey.

The Conditioned Eyelid Response Compared with Conditioned Responses Based on Shock

Turning now to the primary purpose of this study, we find that the lid reflex is much easier to condition in the rat than is the leg or tail withdrawal from shock. It was the exceptional rat that gave more than an occasional conditioned leg flexion. But all four of the rats used in this study developed conditioned eyelid reactions during the first session, and maintained an average frequency in the neighborhood of 60 percent thereafter. The incidence of conditioned lid reflexes is roughly comparable to that reported for conditioned breathing responses in the earlier studies. This was true in spite of conditioning sessions that were probably too long for greatest effectiveness. Further, the mechanical recording system attached to the lid must have been much more distracting than was the one used for recording fore-leg flexion.

It seems clear, then, that fairly localized and specific responses *can* be stably conditioned in the white rat. The difficulty in establishing conditioned fore-leg withdrawal from shock would seem to be due to a combination of restraint and shock, as one of us suggested in an earlier paper.

Summary and Conclusions

1. Conditioned eyelid responses were established and extinguished in each of four rats. The conditioned stimulus, a

buzz, was presented for 800 ms., and terminated simultaneously with the unconditioned stimulus, a puff of air to the cornea.

2. The conditioned response appeared as an anticipatory lid closure in all rats during the first 10 or 20 reinforcements. Its latency decreased as conditioning proceeded, approaching 150–200 ms. as the response became well established.

3. The frequency with which anticipatory conditioned responses occurred reached an average value of 60 percent during the second session, and varied around that value thereafter. Thus it did not become as stable as similar conditioned responses in dog or monkey.

4. The daily records showed a low incidence of conditioned responses at the beginning and end of the session. The early stimuli apparently fell in a 'warming up' period, while the later ones were affected by inhibitory factors which developed as the session continued. The sessions (50 reinforcements in an hour) were too long. This probably accounted for the failure to go beyond a 60 percent average incidence.

5. Extinction was not complete after 2 sessions of 60 unreinforced buzzes each.

6. This experiment shows that a precise and specific response can be conditioned in a white rat in spite of restraint in a holder.

(Manuscript received August 7, 1938)

References

1. Campbell, A. A., and Hilgard, E. R., Individual differences in ease of conditioning, *J. Exper. Psychol.*, 1936, **19**, 561–571.
2. Hilgard, E. R., and Campbell, A. A., The course of acquisition and retention of conditioned eyelid responses in man, *J. Exper. Psychol.*, 1936, **19**, 227–247.
3. Hilgard, E. R., and Marquis, D. G., Acquisition, extinction, and retention of conditioned lid responses to light in dogs, *J. Comp. Psychol.*, 1935, **19**, 29–58.
4. Hilgard, E. R., and Marquis, D. G., Conditioned eyelid responses in monkeys, with a comparison of dog, monkey, and man, *Psychol. Monog.*, 1936, **47**, 186–198.
5. Hovland, C. I., 'Inhibition of reinforcement' and phenomena of experimental extinction, *Proc. Nat. Acad. Sci.*, 1936, **22**, 430–433.
6. Kappauf, W. E., and Schlosberg, H., Conditioned responses in the white rat. III. Conditioning as a function of the length of the period of delay, *J. Genet. Psychol.*, 1937, **50**, 27–45.

7. LIDDELL, H. S., JAMES, W. T., AND ANDERSON, O. D., The comparative physiology of the conditioned motor reflex, *Comp. Psychol. Monog.*, 1934, **11**, No. 1, pp. 89.
8. SCHLOSBERG, H., Conditioned responses in the white rat, *J. Genet. Psychol.*, 1934, **45**, 303–335.
9. ——, Conditioned responses in the white rat. II. Conditioned responses based upon a shock to the foreleg, *J. Genet. Psychol.*, 1936, **49**, 107–138.

ORDERS FOR THE PRESENTATION OF PAIRS IN THE METHOD OF PAIRED COMPARISON

BY ROBERT J. WHERRY

University of North Carolina

This problem is a very old one and the plans suggested have grown successively better. R. T. Ross published a paper in 1934 (**6**) which provides a very good summary of the problem.

After discussing the contributions of Fechner (1871) (**3**), Cohn (1894) (**2**), Kowalewski (1904) (**4**) and Külpe (1907) (**5**), Ross sets forth the general principles involved and presents a new ordering scheme.

The points which Ross stresses as being necessary in any good system are:

(1) elimination of space and time errors,
(2) avoidance of regular repetitions which might influence judgment,
(3) maintenance of the greatest possible spacing between pairs involving any given number of the stimulus group.

To these should be added a fourth, which Ross recognized, but in the solution of which, as we shall see, he violated the first three. This principle is

(4) balancing out of the effects of fatigue.

Ross's article provided a schematic device for giving an *optimum* list of pairs for any length scale. The lists obtained by this method agreed with the empirically derived lists of Kowalewski, recognized by Beebe-Center (**1**) as being the most desirable so far invented. Ross claimed that the order was optimum, his contribution being the extension of the plan to groups of any size (Kowalewski having given the order only for groups of 5, 7, and 15).

The writer does not want to belittle the contribution made. The possibility of obtaining results of this degree of excellence for series of any length is certainly worthwhile. As we shall see, however, these series are not optimum in at least one sense, which will be demonstrated, and one serious error of application was made by Ross, due to his having overlooked a possibility of greatly multiplying the results of his discovery.

Let us study first the claim made for the optimum character of the series in light of the series for seven items. Here Ross gives the preferred series as: 1–2, 7–3, 6–4, 5–1, 3–2, 4–7, 5–6, 1–3, 2–4, 7–5, 6–1, 4–3, 5–2, 6–7, 1–4, 3–5, 2–6, 7–1, 4–5, 3–6, and 2–7.

If these pairs be studied with regard to the reappearance of a given number in the series we find:

TABLE I

ANALYSIS OF ROSS LIST FOR TIME ERRORS

Item	Pairs between reappearance of a given item					
1	2	3	2	3	2	12 = total
2	3	3	3	3	3	15
3	2	2	3	3	3	13
4	2	2	2	2	3	11
5	2	2	2	2	2	10
6	3	3	2	2	2	12
7	3	3	3	3	2	14
						87

If likewise we study the results with respect to space errors (whether the item is presented equally often in both first and second position, and whether these are distributed or massed) we find:

TABLE II

ANALYSIS OF ROSS LIST FOR SPACE ERRORS

Item	Position (1 = first, 2 = second) upon successive appearance *					
1	1	2	1	2	1	2
2	2 (3)	2	1	2	1 (3)	1
3	2	1	2 (3)	2	1 (3)	1
4	2	1	2	1	2	1
5	1 (2)	1	2	1	2 (2)	2
6	1	2	1 (2)	1	2 (2)	2
7	1	2	1	2	1	2

* Figures in parentheses indicate intervening pairs when a time-space error occurs.

It is evident that each item is first just as often as it is second. There are, however, four variables which appear either first, or second, two times in a row; indeed it occurs twice with respect to each of these variables. From this we may conclude that the spacing is not perfectly timed.

While such defects may not be important, they did raise certain doubts in the writer's thinking as to the claim that this order was optimum in the conclusive sense implied.

After a certain amount of trial-and-error the following series was empirically derived: 4–2, 6–7, 3–1, 2–5, 7–4, 1–6, 5–3, 2–7, 4–1, 6–3, 7–5, 1–2, 3–4, 5–6, 7–1, 2–3, 4–6, 1–5, 3–7, 6–2, and 5–4.

If this series be analyzed in the same manner as that of Ross, we have:

TABLE III

ANALYSIS OF WHERRY LIST FOR TIME ERRORS

Item	Pairs between reappearance of a given item					
1	2	2	2	2	2	total = 10
2	2	3	3	3	3	14
3	3	2	2	2	2	11
4	3	3	3	3	3	15
5	2	3	2	3	2	12
6	3	3	3	2	2	13
7	2	2	2	3	3	12
						87

These results are as good, or bad, as those obtained by Ross. No further manipulation succeeded in bettering this result; many other schemes were worse.

With regard to the other method of analysis, a difference appeared. The results were:

TABLE IV

ANALYSIS OF WHERRY LIST FOR SPACE ERRORS *

Item	Position (1 = first, 2 = second) upon successive appearance					
1	2	1	2	1	2	1
2	2	1 (3)	1	2	1	2
3	1	2 (2)	2	1	2	1
4	1	2	1	2	1	2
5	2	1	2	1	2	1
6	1	2	1	2 (2)	2	1
7	2	1	2	1 (3)	1	2

* Figures in parentheses indicate intervening pairs.

While four items are still imperfect, the violation is only once for each of the four; this is just one half of the incidence of this type of error in the Ross optimum list.

In view of this finding, it seems to the writer that the question of optimum lists is still an open one. So far he has not been able to find a rational scheme for bettering all such lists, but surely, if it can be achieved empirically for seven variables (the writer has also bettered the Ross list for five variables in the same manner), it can be done rationally for any number. The search for such a rational method should not be stopped due to a belief that Ross has achieved its solution.

Another, and even more serious, objection to Ross's conclusions occurs in connection with his attempt to solve the problem of fatigue. Ross says, "If the orders are repeated in reverse arrangement directly after the experiment has been completed in the given order, the effects of fatigue may be balanced out. Also such a plan insures perfect balancing for series of an even number of members."

This suggestion, however, immediately leads to violation of all of the other principles so carefully worked out previously. Observe, for example, what happens if we consider the last four items in the series and the immediate repetition of these four items in reverse order in the fatigue elimination series; we have:

1–5, 3–7, 6–2, 5–4; 5–4, 6–2, 3–7, and 1–5.

It is apparent at once that each item is being repeated in the same space order as in the previous list, a violation of our general principles. If we assume that the experimenter notes this fact and inverts the order of presentation within the pairs, we still have:

1–5, 3–7, 6–2, 5–4; 4–5, 2–6, 7–3, and 5–1.

This still leaves us with serious time defects, as indicated in the table below:

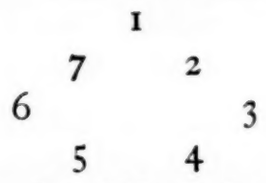
TABLE V

TIME ERRORS EFFECTED BY USE OF FATIGUE SERIES

Item	1	2	3	4	5	6	7
Pairs between reappearance of a given item	6	2	4	0	0	2	4

This in case of items 4 and 5 is a clear violation of our principles. The problem of fatigue effect elimination is important, however. Let us see how it can be solved without violation of the remaining principles involved.

Let us return to the empirically obtained series of the writer, and apply the following idea. It is obvious that the item labeled '1' could be any of the seven items involved, and so on for any of the remaining items. All that we must do is to maintain the same interrelationships among our items



Thus, if all of the items be shifted clockwise one position, we may make the substitution indicated in our ordered series without disturbing any of the relationships involved (note that when seven is increased it becomes 1). This process may be repeated until our original series is reobtained. Doing this we obtain the following series:

TABLE VI

RESULTS FROM ROTATION OF ITEMS IN SERIES

(1)	(2)	(3)	(4)	(5)	(6)	(7)
4–2	5–3	6–4	7–5	1–6	2–7	3–1
6–7	7–1	1–2	2–3	3–4	4–5	5–6
3–1	4–2	5–3	6–4	7–5	1–6	2–7
2–5	3–6	4–7	5–1	6–2	7–3	1–4
7–4	1–5	2–6	3–7	4–1	5–2	6–3
1–6	2–7	3–1	4–2	5–3	6–4	7–5
5–3	6–4	7–5	1–6	2–7	3–1	4–2
2–7	3–1	4–2	5–3	6–4	7–5	1–6
4–1	5–2	6–3	7–4	1–5	2–6	3–7
6–3	7–4	1–5	2–6	3–7	4–1	5–2
7–5	1–6	2–7	3–1	4–2	5–3	6–4
1–2	2–3	3–4	4–5	5–6	6–7	7–1
3–4	4–5	5–6	6–7	7–1	1–2	2–3
5–6	6–7	7–1	1–2	2–3	3–4	4–5
7–1	1–2	2–3	3–4	4–5	5–6	6–7
2–3	3–4	4–5	5–6	6–7	7–1	1–2
4–6	5–7	6–1	7–2	1–3	2–4	3–5
1–5	2–6	3–7	4–1	5–2	6–3	7–4
3–7	4–1	5–2	6–3	7–4	1–5	2–6
6–2	7–3	1–4	2–5	3–6	4–7	5–1
5–4	6–5	7–6	1–7	2–1	3–2	4–3
(8)	(9)	(10)	(11)	(12)	(13)	(14)

It is of course possible to reverse each series either vertically or horizontally, *i.e.*, start at the bottom rather than at the top, or reverse the order within each pair throughout the series. The first of these steps gives rise to series (8) to (14). The second of these steps gives rise to series (1R) through (14R), thus giving us so far a total of twenty-eight series.

It is also possible to reverse completely the original series of numbers without disturbing their interrelationships; thus we may have

Old order, 1234567,
New order, 7654321.

Substituting these quantities in list (1) gives list (15) below. List (15) is then shifted clockwise as was list (1) above, yielding lists (16) through (21); the vertical inverse of each of these is labeled from (22) to (28); and the horizontal inversion of both of these groups are labeled from (15R) through (28R). This brings the total number of lists up to 56. The results are given below:

TABLE VII

FURTHER LISTS DERIVED BY INVERSION AND ROTATION

(15)	(16)	(17)	(18)	(19)	(20)	(21)
4-6	5-7	6-1	7-2	1-3	2-4	3-5
2-1	3-2	4-3	5-4	6-5	7-6	1-7
5-7	6-1	7-2	1-3	2-4	3-5	4-6
6-3	7-4	1-5	2-6	3-7	4-1	5-2
1-4	2-5	3-6	4-7	5-1	6-2	7-3
7-2	1-3	2-4	3-5	4-6	5-7	6-1
3-5	4-6	5-7	6-1	7-2	1-3	2-4
6-1	7-2	1-3	2-4	3-5	4-6	5-7
4-7	5-1	6-2	7-3	1-4	2-5	3-6
2-5	3-6	4-7	5-1	6-2	7-3	1-4
1-3	2-4	3-5	4-6	5-7	6-1	7-2
7-6	1-7	2-1	3-2	4-3	5-4	6-5
5-4	6-5	7-6	1-7	2-1	3-2	4-3
3-2	4-3	5-4	6-5	7-6	1-7	2-1
1-7	2-1	3-2	4-3	5-4	6-5	7-6
6-5	7-6	1-7	2-1	3-2	4-3	5-4
4-2	5-3	6-4	7-5	1-6	2-7	3-1
7-3	1-4	2-5	3-6	4-7	5-1	6-2
5-1	6-2	7-3	1-4	2-5	3-6	4-7
2-6	3-7	4-1	5-2	6-3	7-4	1-5
3-4	4-5	5-6	6-7	7-1	1-2	2-3
(22)	(23)	(24)	(25)	(26)	(27)	(28)

This gives us a total of 56 potential lists, each exactly as good as the original empirically derived list. It is of course possible to apply this same scheme to any of Ross's rationally determined lists with similar increase in their usefulness.

Now if one wishes to avoid the effect of fatigue he may choose some combination from all of these lists. No two lists taken alone will satisfy this condition. It is possible, however, to secure a series of such lists, the joining together of which introduces no errors of any kind, and which as a composite battery will successfully eliminate the fatigue effect.

TABLE VIII

TIME ANALYSIS OF ARTICULATION POINTS

Item	Number of spaces separating each item between lists												
	1 26	26 6R	5R 24R	24R 4	4 22	22 2R	2R 27R	27R 7	7 25	25 5R	5R 23R	23R 3	3 28
1	3	2	3	3	2	2	2	3	3	2	3	3	2
2	3	2	3	3	2	3	3	2	3	3	2	2	2
3	3	3	2	2	2	3	3	2	3	3	2	3	3
4	3	3	2	3	3	2	3	3	2	2	2	3	3
5	2	2	2	3	3	2	3	3	2	3	3	2	3
6	2	3	3	2	3	3	2	2	2	3	3	2	3
7	2	3	3	2	3	3	2	3	3	2	3	3	2

Thus, choosing from our list of 56 potential lists above, we select the following sequence of 14 lists (this includes each of the fourteen basic lists, each put in proper order and sequence to fit the rules) to be adequate: Series I; 1, 26, 6R, 24R, 4, 22, 2R, 27R, 7, 25, 5R, 23R, 3 and 28. (A similar correct sequence would be each of these in the same order but with the R's on the other lists.)

To assure ourselves that no time order is introduced we must investigate the spacing between each of the seven items for the joining of each list with its predecessor. This is reported in the following table.

We have thus shown that other principles are not violated by so combining the 14 series. We have still not demonstrated that this series does, in effect, balance out the fatigue effect. It is obvious that in the 14 lists each made up of 21

pairs, that each pair involved cannot possibly be in each of the twenty one possible positions. This, however, is not necessary. What is necessary is that each one of the twenty one pairs involved shall on the average have been in the 11th position, *i.e.*, that the mean position of each item be eleventh in the series of twenty one. That this is true for this list is shown in Table IX.

TABLE IX

POSITION OF EACH PAIR IN EACH SCALE

Pair	Scale													
	1	26	6R	24R	4	22	2R	27R	7	25	5R	23R	3	28
1–2	12	9	13	10	14	20	15	1	16	6	21	7	2	8
1–3	3	21	7	14	11	11	8	15	1	19	17	16	6	5
1–4	9	13	10	2	18	17	19	18	4	3	5	4	20	12
1–5	18	17	19	18	4	3	5	4	20	12	9	13	10	2
1–6	6	5	3	21	7	14	11	11	8	15	1	19	17	16
1–7	15	1	16	6	21	7	2	8	12	9	13	10	14	20
2–3	16	6	21	7	2	8	12	9	13	10	14	20	15	1
2–4	1	19	17	16	6	5	3	21	7	14	11	11	8	15
2–5	4	3	5	4	20	12	9	13	10	2	18	17	19	18
2–6	20	12	9	13	10	2	18	17	19	18	4	3	5	4
2–7	8	15	1	19	17	16	6	5	3	21	7	14	11	11
3–4	13	10	14	20	15	1	16	6	21	7	2	8	12	9
3–5	7	14	11	11	8	15	1	19	17	16	6	5	3	21
3–6	10	2	18	17	19	18	4	3	5	4	20	12	9	13
3–7	19	18	4	3	5	4	20	12	9	13	10	2	18	17
4–5	21	7	2	8	12	9	13	10	14	20	15	1	16	6
4–6	17	16	6	5	3	21	7	14	11	11	8	15	1	19
4–7	5	4	20	12	9	13	10	2	18	17	19	18	4	3
5–6	14	20	15	1	16	6	21	7	2	8	12	9	13	10
5–7	11	11	8	15	1	19	17	16	6	5	3	21	7	14
6–7	2	8	12	9	13	10	14	20	15	1	16	6	21	7

If each of these rows is added and averaged, the mean position of each item is indeed found to be 11th in the scale. The use of this entire series would, therefore, adequately cancel out the fatigue effect introduced by the use of a single list.

Since, however, one of the main objections raised against the method of paired comparisons is its length when the number of items becomes at all large, this suggestion of having each individual go through 14 lists (in general $2n$ lists, where n is the number of items) will most certainly be unwelcome.

It is still possible, however, to utilize this finding. If the number of subjects used is always taken as some multiple of $2n$ (here 14, 28, 42, 56, 70, etc.), it will be possible to use each of the selected lists an equal number of times, even though each subject is asked to react to a single list, thus securing a balancing out of the fatigue factor.

Conclusions

1. It is shown that Ross's optimum lists are not optimum in all senses, thus reopening the search for a rational solution of this problem.

2. An empirically derived list for seven items is presented which is superior to the list given by Ross. (Note: the superior list for five items which was referred to above, but not given is 1–2, 3–4, 5–1, 2–3, 4–5, 3–1, 4–2, 5–3, 1–4, and 2–5. This list has two repetitional space errors, while the Ross list has four.)

3. A method is given, whereby any list, arrived at either rationally or empirically, may be rewritten in $8n$ different ways, by use of the following steps:

(*a*) Inverse code substitution of the terms; *e.g.*, 1 for n, 2 for $n - 1$, 3 for $n - 2$, etc. This gives us 2 lists.

(*b*) Successive increases of one for each item in each of these lists, the process being repeated until the next increase would give the original list again (when the number n is *raised* it becomes one). This gives a total of $2n$ lists.

(*c*) These lists may be administered from either end. This gives a total of $4n$ lists.

(*d*) Each of the lists may be read either from left-to-right or right-from-left. This makes $8n$ lists.

4. It is shown that $2n$ of these lists may be combined in such a fashion that fatigue effects, introduced by the use of a single list, are cancelled out.

5. It is suggested as an alternative to the above that the number of subjects he selected as some multiple of $2n$, and that each list be used in its proper order an equal number of times, each subject doing just one list.

(Manuscript received June 27, 1938)

References

1. Beebe-Center, J. C., *The Psychology of Pleasantness and Unpleasantness*, Van Nostrand Co., 1932, p. 32.
2. Cohn, J., Experimentelle Untersuchen uber die Gefuhlsbetonung der Farben, Helligkeiten und Ihrer Combination, *Philos. Studien*, 1894, **10**, 562 ff.
3. Fechner, G. T., Zur Experimentellen Aesthetic, *Abh. d. Math.-Phys. Cl. d. k. Schs. Ges. d. Wissensch.*, **9**, Leipzig, 1871, 630.
4. Kowalewski, A., Studien zur Psychologie des Pessimismus, *Grenzfragen des Nerven- und Seelenlebens*, 1904, **24**, 69 ff.
5. Külpe, O., Der Gegenwartig Stand der Experimentellen Aesthetic, *Bericht u. d. Kongr. f. exper. Psychol.*, 1907, **11**.
6. Ross, R. T., Optimum orders for the presentation of pairs in the method of paired comparisons, *J. Educ. Psychol.*, 1934, **25**, 375–382.

AN ELECTRICALLY CONTROLLED FILM SLIDE PROJECTOR [1]

BY JACK BUEL

Wesleyan University

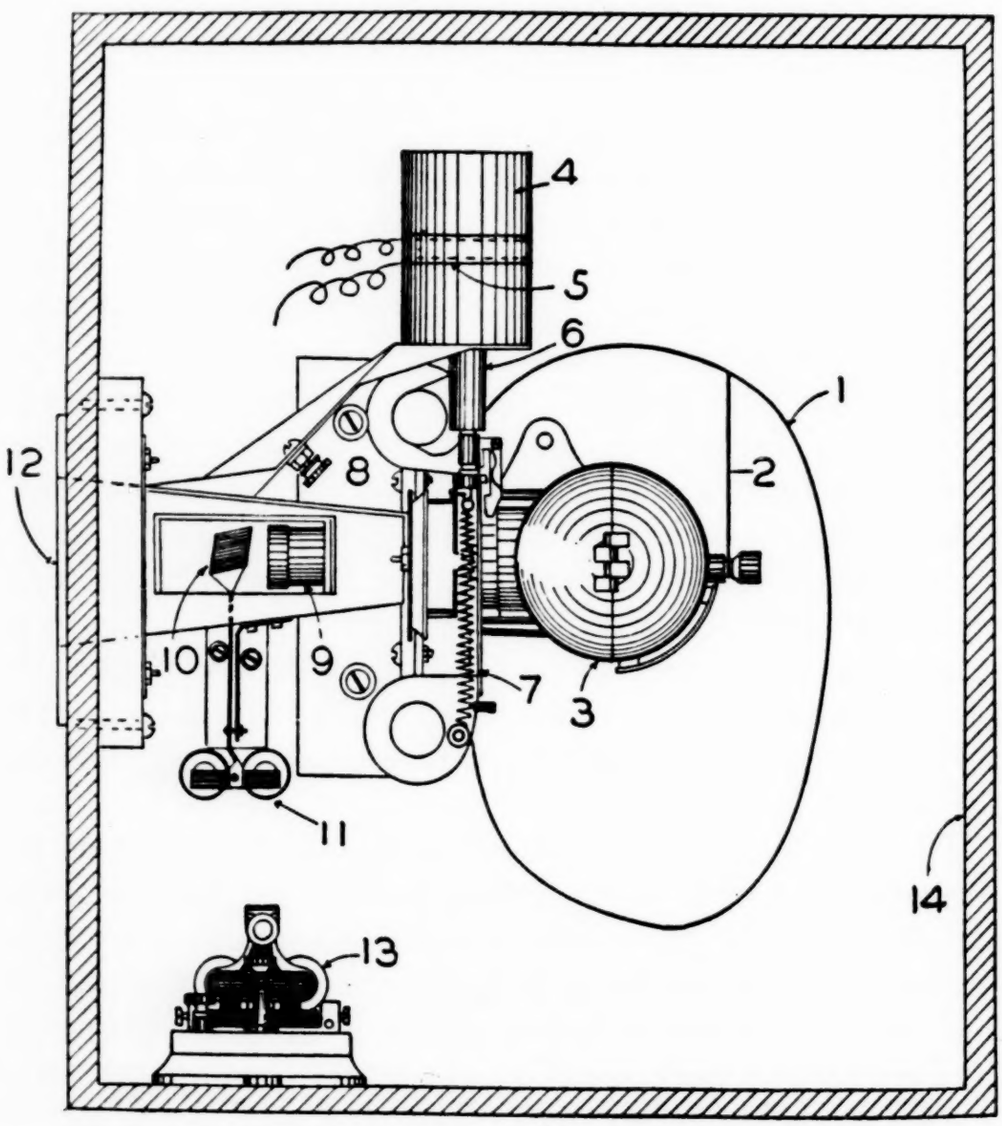
The projector described below will be found useful in many experimental situations and may also be used with advantage in demonstrating memory and tachistoscopic methods in the classroom. The apparatus is a projector of stills from 35 mm film, but has the advantages of being self-enclosed and noiseless, electrically operated and remotely controlled by either subject or experimenter. It provides for the serial presentation of any amount of material for determined exposure intervals and provides with a suitable accessory, for recording both the time at which the exposure took place and its duration. Some further advantages are its portability, easy construction and low cost.

The apparatus is constructed around a Brayco 35 mm projector, mounted upside down in an enclosing cabinet. The base of this projector originally attached to the lamp housing (3) has been removed.

To the front of the projector has been added a metal tube $4\frac{1}{2}$ inches long with a supporting bracket (8) for the frame changing solenoid (4). This tube has flanges at both ends, those at the wider end being bolted to the walls of the cabinet, while the narrow end is bolted to the front of the projector. The tube has the dual role of a support for the apparatus and the two solenoids (4 and 11). Into the side of the tube a window has been cut to allow the objective lens (9) to be adjusted for focus, and for inspection of the magnetically operated shutter vane (10) which is used to determine the exposure.

[1] The general plan for this apparatus was suggested by Dr. Harold E. Jones, Director of the Institute of Child Welfare, University of California. It was designed for use in connection with an experimental cabinet for the study of the emotions, described by Jones (1).

The solenoid (4) has a movable core (6) which is attached to a lever. When a current is passed through the solenoid the core is drawn vertically upward into the center of its magnetic field, and the attached lever advances the film by one frame. It is now apparent that the reason for mounting the projector



SIDE VIEW DIAGRAM OF THE ELECTRICALLY CONTROLLED FILM SLIDE PROJECTOR

1, Endless strip of 33 mm moving picture film. 2, Wire support to hold film away from lamp housing. 3, Lamp housing. 4, Frame changing solenoid coil. 5, Auxiliary coil to supply induced current to recording galvanometer. 6, Iron core plunger or solenoid. 7, Extension spring to return frame changing lever to original position after a frame change. 8, Support tube and bracket for operating solenoid. 9, Objective lens focusing tube. 10, Shutter vane controlling exposure. 11, Double wire wound magnets activating shutter. 12, Glass projection window. 13, Heavy contact relay controlling operating solenoid. 14, Sound proof covering box.

upside down is so that the weight of the moveable solenoid core will return the lever to its original position after the new frame has been brought into place. The extension spring (7) insures the positive return of the lever. The outside dimensions of the solenoid are $2\frac{1}{8}$ inches by $3\frac{1}{4}$ inches wound full with number 18 D.C.C. magnet wire. The movable core is 9/16 inches and 3 inches and is fastened directly to the frame changing lever through rubber washers. The washers reduce the shock of the rapid core movement. A further modification of the projector is in the removal of the rewind ratchet in order to reduce friction. This is not a disadvantage since a continuous strip film is more applicable to psychological work than a cartridge roll of film.

The exposure shutter (10) is motivated by magnets taken from a standard door bell, and is mounted on a bracket fastened to the metal support tube.

Around the operating solenoid (4) may be seen several turns of wire (5). This coil serves to motivate a galvanometer indicator at the recording polygraph. The induced current indicates exactly when a new frame was introduced and for how long.

The operating solenoid requires 25 volts, but it may be constructed to other voltage requirements. In order to reduce switch contact difficulties the solenoid is controlled through a 4 ohm 6 volt telegraph relay (13) fitted with heavy silver contacts.

In practice it has been found that the 10 volt lamp and resistance cord with which the projector is supplied does not give a sufficiently bright image. An 8 volt lamp substituted for the 10 volt one will give the proper illumination and will last very well under the conditions usually obtaining for such apparatus. A white cardboard screen about seven feet from the projector gives a brilliant image 21 inches by 19 inches.

By connecting a pendent hand push button switch to the relay the subject may himself control the frequency and duration of exposures or the experimenter may control the entire operation, and by means of the shutter obtain standard exposures. In the particular group of experiments for which

this apparatus was designed, the pictures used comprised a variety of emotionally provocative material; the length of exposure of each picture was recorded on a photo-polygraph together with records of electrodermal responses, pulse rate and other somatic and visceral processes (2).

Since the whole apparatus, including the sound proof box is but 12 inches by 15 inches by 18 inches it readily lends itself to classroom demonstrations as well as to more exacting experimental requirements.

(Manuscript received July 6, 1938)

REFERENCE

1. Jones, H. E. An experimental cabinet for physiological studies of emotions. *Child Development*, 1936, 7, 183–188.